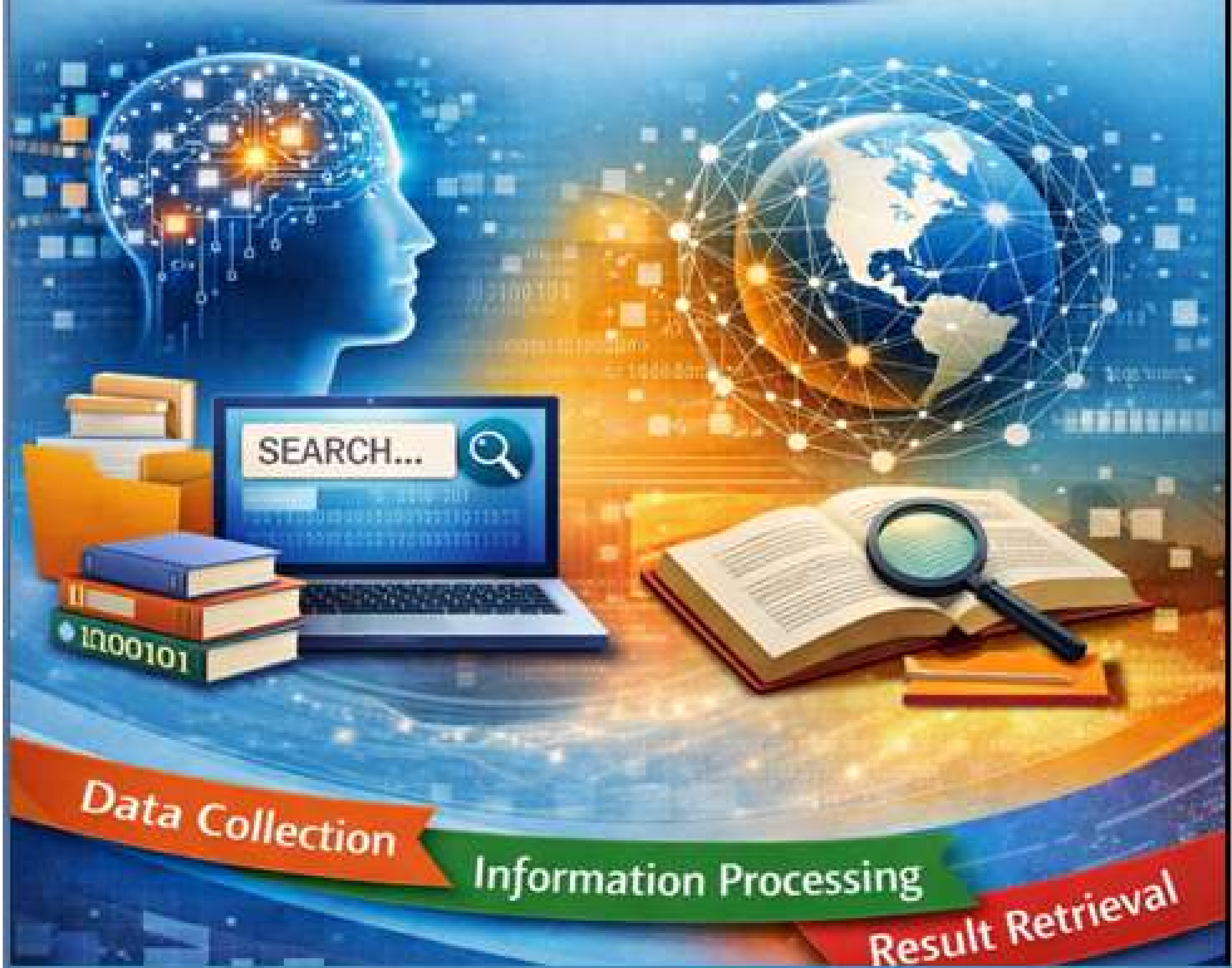
MLIS-202

सूचना प्रसंस्करण और पुनर्प्राप्ति

Information Processing and Retrieval

SEMESTER-II



उत्तराखण्ड मुक्त विश्वविद्यालय

हल्द्वानी

पुस्तकालय एवं सूचना विज्ञान विद्याशाखा

पाठ्यक्रम समिति

प्रो जयदीप शर्मा
सामाजिक विज्ञान विद्यापीठ,
इंदिरा गांधी राष्ट्रीय मुक्त विश्वविद्यालय,
नई दिल्ली।

प्रो .मनोज कुमार जोशी
पुस्तकालय एवं सूचना विज्ञान विभाग
कुरुक्षेत्र विश्वविद्यालय कुरुक्षेत्र हरियाणा ।

डॉ. शत्रुघन झा
पुस्तकालयध्यक्ष
गुरुकुल कांगड़ी विश्वविद्यालय हरिद्वार।

प्रो.अरविंद भट्ट.
विभागाध्यक्ष पुस्तकालय एवं सूचना विज्ञान
विद्याशाखा उत्तराखंड मुक्त विश्विद्यालय,
हल्द्वानी।

प्रीति शर्मा
असिस्टेन्ट प्रोफेसर (ए0 सी0) एवं
कार्यक्रम समन्वयक
पुस्तकालय एवं सूचना विज्ञान विभाग
उत्तराखण्ड मुक्त विश्वविद्यालय, हल्द्वानी

पाठ्यक्रम समन्वयक एवं संयोजन	सम्पादन
सुश्री प्रीति शर्मा असिस्टेन्ट प्रोफेसर एवं कार्यक्रम समन्वयक पुस्तकालय एवं सूचना विज्ञान विद्याशाखा उत्तराखण्ड मुक्त विश्वविद्यालय, हल्द्वानी।	सुश्री प्रीति शर्मा पुस्तकालय एवं सूचना विज्ञान विद्याशाखा उत्तराखण्ड मुक्त विश्वविद्यालय, हल्द्वानी।

इकाई लेखन	खण्ड	इकाई संख्या
सुश्री प्रीति शर्मा असिस्टेन्ट प्रोफेसर, पुस्तकालय एवं सूचना विज्ञान विभाग पुस्तकालय एवं सूचना विज्ञान विद्याशाखा उत्तराखण्ड मुक्त विश्वविद्यालय, हल्द्वानी।	1,2,3,4	1 से 14 इकाई

कापीराइट @उत्तराखण्ड मुक्त विश्वविद्यालय

प्रकाशन वर्ष – 2025 प्रकाशक- उत्तराखण्ड मुक्त विश्वविद्यालय, हल्द्वानी।

मुद्रक: -

नोट : - (इस पुस्तक के समस्त इकाईयों के लेखन तथा कॉपीराइट संबंधी किसी भी मामले के लिये संबंधित इकाई लेखक जिम्मेदार होगा। किसी भी विवाद का निस्तारण नैनीताल स्थित उच्च न्यायालय अथवा हल्द्वानी सत्रीय न्यायालय में किया जायेगा।)

अनुक्रम

प्रथम खण्ड – सूचना का संगठन **पृष्ठ - 2**

इकाई 1: सूचना का बौद्धिक संगठन 3-20
इकाई 2: अनुक्रमण भाषाएँ : अवधारणाएँ और प्रकार, विषय शीर्षकसूचियाँ एवं थिसॉरस 21- 42
इकाई 3: अनुक्रमण भाषाएँ : वर्गीकरण योजना 43-59
इकाई 4: अनुक्रमण प्रणाली और तकनीक 60-77
इकाई 5: अनुक्रमण प्रणाली का मूल्यांकन 78-96

द्वितीय खण्ड – ग्रंथ सूची विवरण **पृष्ठ 97**

इकाई 6: ग्रंथ सूची विवरण के सिद्धांत और विकास 98-123
इकाई 7: ग्रंथ सूची विवरण के नियम 124-143
इकाई 8: ग्रंथ सूची रिकॉर्ड प्रारूप के लिए मानक 144-163

तृतीय खण्ड - सूचना भंडारण और पुनर्प्राप्ति प्रणाली **पृष्ठ 164**

इकाई 9: मेटाडेटा: MARC 21-856 फ़ील्ड, डबलिनकोर 165-192
इकाई 10: ISAR सिस्टम: उद्देश्य और प्रकार, संचालन और डिज़ाइन 193-211
इकाई 11: ISAR सिस्टम की अनुकूलता 212-237

चतुर्थ खण्ड - सूचना पुनर्प्राप्ति **पृष्ठ 238**

इकाई 12: इंटेलिजेंट IR सिस्टम 239-253
इकाई 13: सूचना पुनर्प्राप्ति - प्रक्रियाएँ और तकनीक 254-275
इकाई 14: सूचना पुनर्प्राप्ति-मॉडल और उनके अनुप्रयोग 276-298

M.L.I.S. -25

द्वितीय सेमेस्टर

सूचना प्रसंस्करण और पुनर्प्राप्ति

(Information Processing and Retrieval)

(MLIS-202)

प्रथम खण्ड

सूचना का संगठन
(Organization of Information)

इकाई–1 : सूचना का बौद्धिक संगठन (Intellectual Organization of Information)

इकाई की रूपरेखा

1.1 प्रस्तावना

1.2 इकाई के उद्देश्य

1.3 सूचना संगठन: अर्थ एवं अवधारणा

1.4 बौद्धिक संगठन: परिभाषाएँ एवं दृष्टिकोण

1.5 सूचना/ज्ञान संगठन का ऐतिहासिक विकास

1.6 सूचना संगठन का मूल्य एवं महत्व

1.7 बौद्धिक संगठन की विशेषताएँ

1.8 सूचना खोजना एवं उपयोग: संगठन की भूमिका

1.9 सूचना आवश्यकता एवं संगठन का संबंध

1.10 सूचना आवश्यकता के मॉडल (Conceptual Models)

1.11 सूचना संसाधनों/स्रोतों का चयन: मानदंड

1.12 सूचना अधिभार एवं संगठन की उपयोगिता

1.13 सूचना स्रोत बनाम सूचना संसाधन

1.14 सूचना स्रोतों की विशेषताएँ

1.15 सारांश

1.16 शब्दावली

1.17 अभ्यास प्रश्नों के संकेत

1.18 संदर्भ ग्रंथ सूची

1.19 निबंधात्मक प्रश्न

1.1 प्रस्तावना (Introduction)

मानव सभ्यता के विकास में सूचना (Information) की भूमिका अत्यंत महत्वपूर्ण रही है। जैसे-जैसे समाज जटिल होता गया, वैसे-वैसे सूचना की मात्रा, स्वरूप और उपयोग में तीव्र वृद्धि हुई। आज का युग **सूचना युग (Information Age)** कहलाता है, क्योंकि शिक्षा, शोध, प्रशासन, उद्योग, विज्ञान, प्रौद्योगिकी और नीति-निर्माण—सभी क्षेत्रों में सूचना के बिना कार्य की कल्पना भी नहीं की जा सकती।

हालाँकि सूचना की प्रचुरता अपने आप में उपयोगी नहीं होती। यदि सूचना को **उचित ढंग से व्यवस्थित (organized)** न किया जाए, तो वह उपयोगकर्ता के लिए भ्रम, समय की बर्बादी और निर्णय में कठिनाई उत्पन्न करती है। इसी समस्या के समाधान के रूप में **सूचना का बौद्धिक संगठन (Intellectual Organization of Information)** की अवधारणा सामने आती है।

सूचना का बौद्धिक संगठन वह प्रक्रिया है जिसके अंतर्गत सूचना को उसके **विषय, अर्थ, संदर्भ और उपयोगकर्ता की आवश्यकता** के अनुसार तार्किक एवं वैज्ञानिक ढंग से संरचित किया जाता है। यह केवल भौतिक व्यवस्था (जैसे पुस्तकों को शेल्फ पर रखना) नहीं है, बल्कि एक **मानसिक एवं विश्लेषणात्मक प्रक्रिया** है, जिसमें ज्ञान की संरचना पर विशेष ध्यान दिया जाता है।

1.2 इकाई के उद्देश्य

इस इकाई के अध्ययन के पश्चात शिक्षार्थी:

- सूचना एवं बौद्धिक संगठन की मूल अवधारणा को स्पष्ट रूप से समझ सकेंगे
- बौद्धिक संगठन के महत्व और आवश्यकता का विश्लेषण कर सकेंगे
- सूचना खोज, उपयोग एवं आवश्यकता में संगठन की भूमिका को समझ सकेंगे
- सूचना अधिभार की समस्या और संगठन के समाधानात्मक पक्ष को पहचान सकेंगे
- सूचना स्रोतों एवं संसाधनों का आलोचनात्मक मूल्यांकन कर सकेंगे
- आधुनिक पुस्तकालय एवं डिजिटल परिवेश में सूचना संगठन की भूमिका को समझ सकेंगे

1.3 सूचना संगठन: अर्थ एवं अवधारणा

सूचना संगठन (Information Organization) का अर्थ है—सूचना को इस प्रकार व्यवस्थित करना कि वह **समझने योग्य (understandable)**, **खोजने योग्य (searchable)** और **उपयोग योग्य (usable)** बन सके।

यह संगठन केवल भौतिक (physical) नहीं होता, बल्कि मुख्यतः **बौद्धिक (intellectual)** होता है, जिसमें विषय विश्लेषण, वर्गीकरण, अनुक्रमण और वर्णन शामिल होते हैं।

उदाहरण

यदि किसी पुस्तकालय में 10,000 पुस्तकें बिना किसी व्यवस्था के रख दी जाएँ, तो वे व्यावहारिक रूप से **बेकार** हो जाएँगी।
लेकिन जब उन्हें **DDC/UDC**, विषय शीर्षकों और अनुक्रमण के माध्यम से व्यवस्थित किया जाता है, तब वही संग्रह **ज्ञान का सशक्त माध्यम** बन जाता है।

1.3 सूचना संगठन: अर्थ एवं अवधारणा (Meaning and Concept of Information Organization)

सूचना संगठन का अर्थ है—
सूचना को इस प्रकार व्यवस्थित करना कि वह **पहचाने योग्य (identifiable)**,
खोजने योग्य (retrievable) और
उपयोगी (usable) बन सके।

सूचना संगठन में निम्नलिखित बौद्धिक क्रियाएँ सम्मिलित होती हैं:

- विषय विश्लेषण (Subject Analysis)
- अवधारणाओं की पहचान (Concept Identification)
- वर्गीकरण (Classification)
- अनुक्रमण (Indexing)
- वर्णनात्मक विवरण (Bibliographic Description)

उदाहरण

यदि किसी शोधार्थी को "Digital Preservation" पर अध्ययन करना है और पुस्तकालय में सामग्री अव्यवस्थित है, तो वह सही सूचना तक नहीं पहुँच पाएगा। लेकिन यदि वही सामग्री विषय, उपविषय और की-वर्ड के आधार पर संगठित हो, तो सूचना संगठन सफल माना जाएगा।

(Meaning and Concept of Information Organization)

सूचना संगठन (Information Organization) का तात्पर्य उस प्रक्रिया से है जिसके माध्यम से सूचना को इस प्रकार व्यवस्थित किया जाता है कि वह उपयोगकर्ता के लिए आसानी से समझने योग्य, खोजने योग्य तथा उपयोगी बन सके। आधुनिक पुस्तकालय एवं सूचना

प्रणालियों में सूचना की मात्रा निरंतर बढ़ रही है, इसलिए सूचना को केवल संग्रहित करना पर्याप्त नहीं है, बल्कि उसे सुव्यवस्थित रूप में प्रस्तुत करना भी आवश्यक हो गया है।

सूचना संगठन मुख्यतः एक **बौद्धिक (intellectual) प्रक्रिया** है। यह केवल पुस्तकों या दस्तावेज़ों को अलमारियों में रखने तक सीमित नहीं होती, बल्कि इसमें सूचना की विषयवस्तु को समझना, उसका विश्लेषण करना और उसे उपयुक्त ढाँचे में प्रस्तुत करना शामिल होता है। इस प्रक्रिया के अंतर्गत निम्नलिखित प्रमुख क्रियाएँ सम्मिलित होती हैं:

- विषय विश्लेषण (Subject Analysis), जिसके माध्यम से दस्तावेज़ के मुख्य विषय की पहचान की जाती है।
- अवधारणाओं की पहचान (Concept Identification), ताकि विषय से जुड़े प्रमुख विचार स्पष्ट हो सकें।
- वर्गीकरण (Classification), जिससे समान विषयों वाले संसाधनों को एक साथ रखा जा सके।
- अनुक्रमण (Indexing), ताकि विषयों तक शीघ्र और सटीक पहुँच संभव हो।
- वर्णनात्मक विवरण (Bibliographic Description), जिससे संसाधन की पहचान और उपयोग आसान हो सके।

सूचना संगठन का महत्व व्यावहारिक उदाहरण से आसानी से समझा जा सकता है। यदि किसी पुस्तकालय में बड़ी संख्या में पुस्तकें और अन्य सूचना संसाधन बिना किसी व्यवस्था के रख दिए जाएँ, तो वे उपयोगकर्ता के लिए लगभग अनुपयोगी हो जाते हैं। इसके विपरीत, जब वही संसाधन उपयुक्त वर्गीकरण योजना (जैसे DDC या UDC), विषय शीर्षकों और अनुक्रमण तकनीकों के माध्यम से व्यवस्थित किए जाते हैं, तो उपयोगकर्ता अपनी आवश्यक सूचना तक शीघ्रता से पहुँच सकता है।

इसी प्रकार, यदि किसी शोधार्थी को "Digital Preservation" विषय पर अध्ययन करना हो और संबंधित सामग्री विषय, उपविषय तथा की-वर्ड के आधार पर संगठित हो, तो सूचना की खोज सरल और प्रभावी हो जाती है। इस स्थिति में सूचना संगठन को सफल माना जाता है, क्योंकि वह उपयोगकर्ता की सूचना आवश्यकता को संतुष्ट करता है।

1.4 बौद्धिक संगठन की परिभाषाएँ एवं दृष्टिकोण (Definitions and Approaches):

बौद्धिक संगठन वह वैचारिक ढाँचा प्रदान करता है जिसके अंतर्गत सूचना के ज्ञानात्मक तत्वों को व्यवस्थित किया जाता है। विभिन्न विद्वानों ने इसे ज्ञान की संरचना को दर्शाने वाली प्रक्रिया के रूप में देखा है। कुछ दृष्टिकोण विषय को केंद्र में रखते हैं, जबकि कुछ उपयोगकर्ता

की आवश्यकता को प्राथमिकता देते हैं। आधुनिक दृष्टिकोण में ज्ञान संगठन और सूचना संगठन को परस्पर पूरक माना जाता है, जहाँ सूचना का संगठन ज्ञान तक पहुँच का माध्यम बनता है।

"**बौद्धिक संगठन** वह प्रक्रिया है जिसमें सूचना के **विषय, अर्थ और संदर्भ** का विश्लेषण किया जाता है।"

बौद्धिक संगठन का तात्पर्य उस प्रक्रिया से है जिसमें सूचना के **ज्ञानात्मक तत्वों (knowledge elements)** को व्यवस्थित किया जाता है।

प्रमुख दृष्टिकोण:

1. **विषयात्मक दृष्टिकोण (Subject-based Approach)**
2. **उपयोगकर्ता-केंद्रित दृष्टिकोण (User-centred Approach)**
3. **ज्ञान संगठन दृष्टिकोण (Knowledge Organization Approach)**

बौद्धिक संगठन सूचना को केवल संग्रह नहीं, बल्कि **ज्ञान संरचना का प्रतिबिंब** बनाता है।

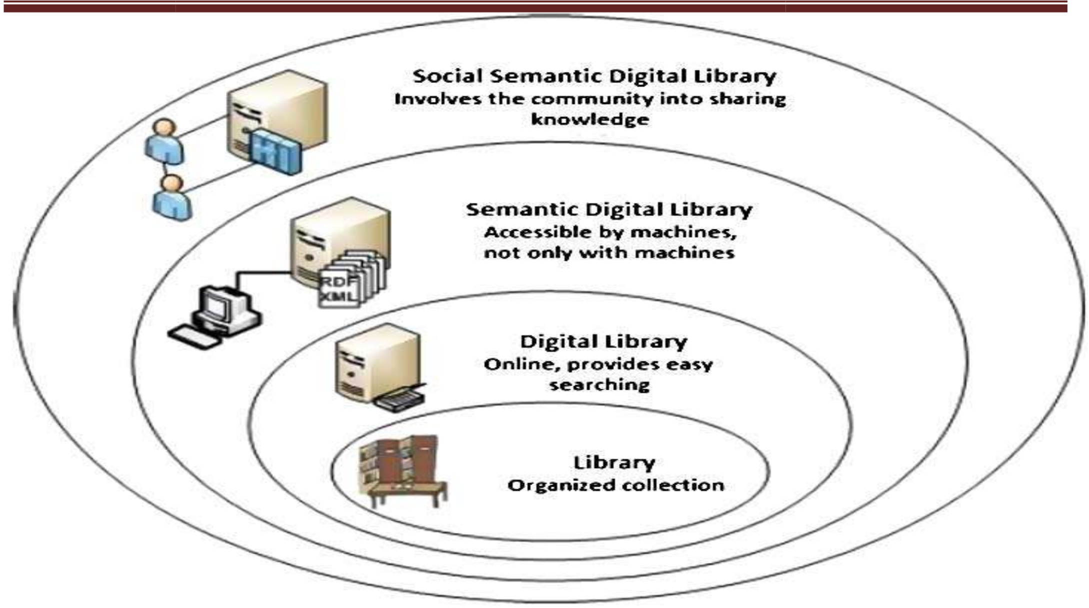
1.5 सूचना/ज्ञान संगठन का ऐतिहासिक विकास (Historical Evolution):

सूचना संगठन का विकास मानव सभ्यता के विकास के साथ जुड़ा हुआ है। प्रारंभिक समाजों में मौखिक परंपरा के माध्यम से ज्ञान का संचार होता था। लेखन और मुद्रण के विकास के साथ सूचना को संग्रहित और व्यवस्थित करने की आवश्यकता बढ़ी। पुस्तकालयों के उद्भव के साथ वर्गीकरण और सूचीकरण की प्रणालियाँ विकसित हुईं। बीसवीं शताब्दी में कंप्यूटर और सूचना प्रौद्योगिकी के आगमन ने सूचना संगठन को एक नया आयाम दिया, जहाँ डिजिटल डेटाबेस, मेटाडाटा और सेमान्टिक वेब जैसी अवधारणाएँ सामने आईं।

सूचना संगठन समय के साथ विकसित हुआ है—मौखिक परंपरा से लेकर डिजिटल और सेमान्टिक वेब तक।

- मौखिक परंपरा (Oral Tradition)
- हस्तलिखित ग्रंथ
- मुद्रण युग
- पुस्तकालय वर्गीकरण योजनाएँ
- कंप्यूटर आधारित कैटलॉग, OPAC, डेटाबेस
- डिजिटल पुस्तकालय, मेटाडाटा,(Metadata, Ontologies)सेमान्टिक वेब

यह विकास दर्शाता है कि सूचना संगठन एक **गतिशील एवं विकासशील प्रक्रिया** है।



1.6 सूचना संगठन का मूल्य एवं महत्व (Value and Significance):

सूचना संगठन का मूल्य इस बात से आँका जाता है कि वह उपयोगकर्ता के लिए सूचना को कितना **प्रभावी और सुलभ** बनाता है। सूचना संगठन का मूल्य इस बात से स्पष्ट होता है कि यह सूचना को कितनी प्रभावी ढंग से उपयोगकर्ता तक पहुँचाता है। सुव्यवस्थित सूचना न केवल समय और श्रम की बचत करती है, बल्कि शोध की गुणवत्ता में भी वृद्धि करती है। निर्णय-निर्माण की प्रक्रिया में संगठित सूचना विश्वसनीय आधार प्रदान करती है। डिजिटल युग में, जहाँ सूचना की मात्रा अत्यधिक है, संगठन सूचना अधिभार की समस्या को नियंत्रित करने में सहायक होता है।

सूचना संगठन का मूल्य निम्न बिंदुओं से स्पष्ट होता है:

- सूचना की त्वरित प्राप्ति
- सूचना की प्रासंगिकता (**relevance**) में वृद्धि
- शोध कार्य में समय और श्रम की बचत
- सूचना अधिभार की समस्या का समाधान
- ज्ञान का प्रभावी प्रसार

उदाहरण: एक सुव्यवस्थित डिजिटल रिपॉजिटरी शोधकर्ताओं के लिए वैश्विक स्तर पर ज्ञान का द्वार खोल देती है।

1.7 बौद्धिक संगठन की विशेषताएँ(Characteristics): बौद्धिक संगठन की विशेषताएँ उसे साधारण व्यवस्था से अलग बनाती हैं। यह प्रक्रिया तार्किक होती है और विषयगत संबंधों पर आधारित होती है। मानकीकरण के माध्यम से इसमें एकरूपता सुनिश्चित की जाती है, जिससे संसाधन-साझाकरण संभव हो पाता है। सबसे महत्वपूर्ण विशेषता यह है कि बौद्धिक संगठन उपयोगकर्ता-केंद्रित होता है और बदलती आवश्यकताओं के अनुसार स्वयं को अनुकूलित करता है।

बौद्धिक संगठन कुछ विशिष्ट गुणों के कारण साधारण व्यवस्था से अलग होता है।

- तार्किक (Logical)
- विषय-आधारित (Subject-oriented)
- मानकीकृत (Standardized)
- उपयोगकर्ता-केंद्रित (User-oriented)
- गतिशील (Dynamic)
- पुनः उपयोग योग्य (Reusable)

1.8 सूचना खोजना एवं उपयोग: संगठन की भूमिका (Role in Information Seeking and Use):

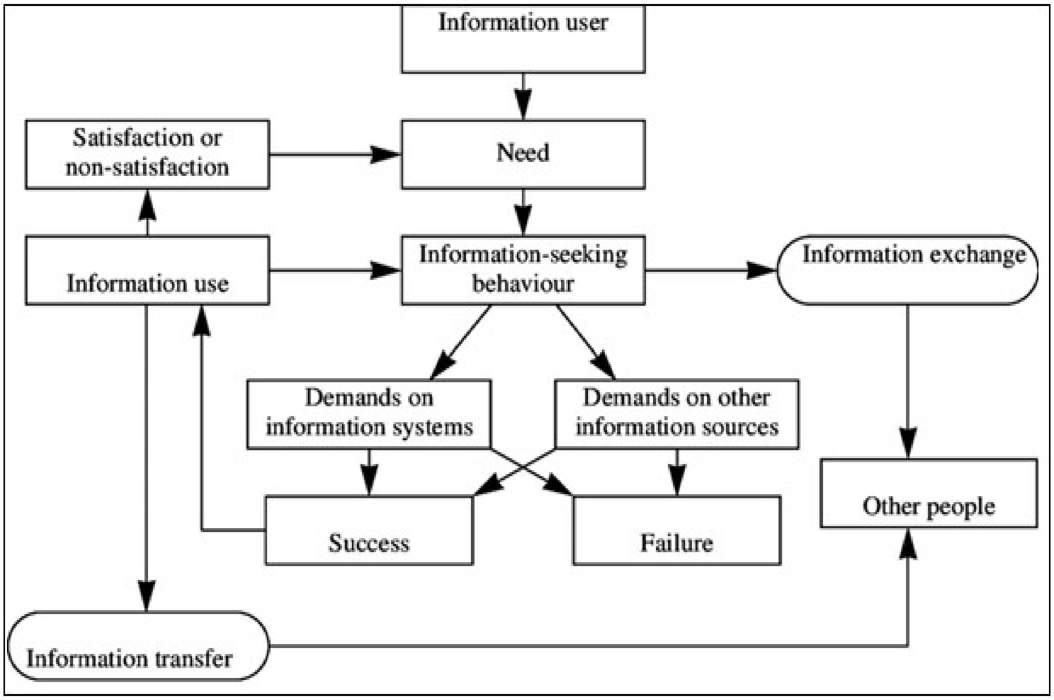
सूचना खोज एक बौद्धिक (cognitive) प्रक्रिया है, जिसमें उपयोगकर्ता अपनी आवश्यकता के अनुसार सूचना की पहचान, खोज और उपयोग करता है। बौद्धिक संगठन इस प्रक्रिया को सरल, सटीक और प्रभावी बनाता है।

बौद्धिक संगठन:

- खोज को सरल बनाता है
- सटीक परिणाम देता है
- उपयोगकर्ता की संतुष्टि बढ़ाता है

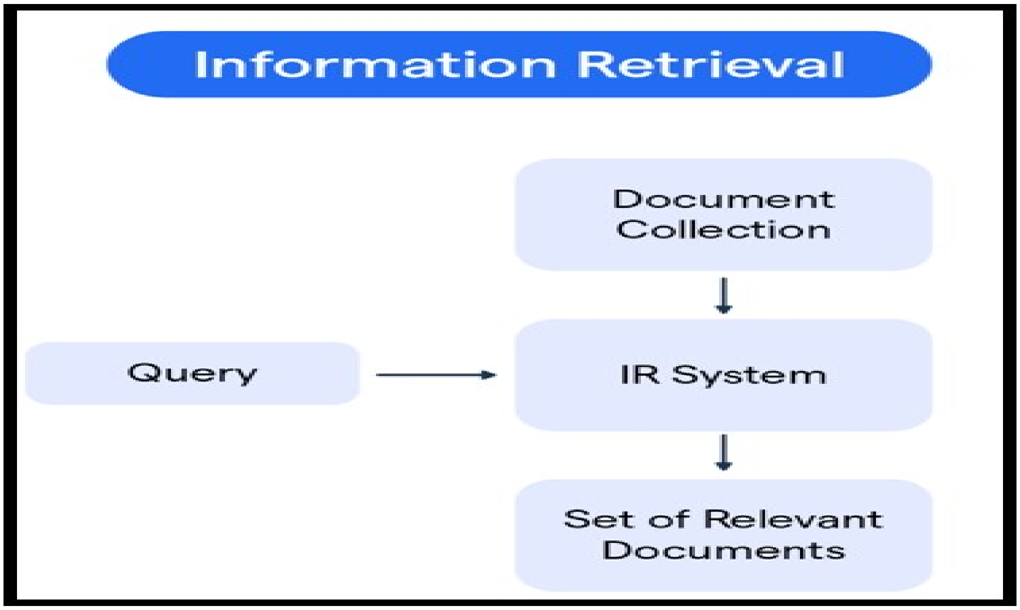
उदाहरण

यदि OPAC या डिजिटल डेटाबेस में उचित अनुक्रमण न हो, तो खोज असफल हो जाती है।



(Fig. 1: Information Seeking Process Model – showing relation between user, need, search, use and satisfaction)

यह चित्र यह दर्शाता है कि किस प्रकार सूचना संगठन उपयोगकर्ता की आवश्यकता और सूचना उपयोग के बीच सेतु का कार्य करता है।



Explanation: यह चित्र यह दर्शाता है कि **सूचना पुनर्प्राप्ति प्रणाली (Information Retrieval System)** किस प्रकार उपयोगकर्ता की **Query** के आधार पर उपयुक्त दस्तावेज़ खोजकर प्रस्तुत करती है।

1.9 सूचना आवश्यकता एवं संगठन का संबंध (Information Need and Organization): सूचना आवश्यकता उपयोगकर्ता के ज्ञान में विद्यमान अंतर से उत्पन्न होती है। यह अंतर तभी भर सकता है जब सूचना उपयुक्त ढंग से संगठित हो। बौद्धिक संगठन सूचना आवश्यकता और सूचना प्राप्ति के बीच सेतु का कार्य करता है, जिससे उपयोगकर्ता की शैक्षणिक और शोध संबंधी आवश्यकताएँ पूर्ण होती हैं।

सूचना आवश्यकता वह **अंतर (gap)** है जो उपयोगकर्ता के वर्तमान ज्ञान और वांछित ज्ञान के बीच होता है। सूचना संगठन इस अंतर को **कम करने का माध्यम** है।

"सूचना आवश्यकता वह स्थिति है जब उपयोगकर्ता अपने ज्ञान में किसी **रिक्तता (gap)** को महसूस करता है।
सूचना संगठन इस रिक्तता को भरने का माध्यम बनता है।"

बिंदुवार विस्तार:

- आवश्यकता की पहचान
- उपयुक्त संसाधन तक पहुँच
- ज्ञान निर्माण में सहायता

1.10 सूचना आवश्यकता के वैचारिक मॉडल (Conceptual Models of Information Need)

सूचना आवश्यकता को समझने के लिए विद्वानों ने विभिन्न **वैचारिक मॉडल (Conceptual Models)** प्रस्तुत किए हैं। ये मॉडल यह स्पष्ट करते हैं कि **उपयोगकर्ता के मन में सूचना की आवश्यकता कैसे उत्पन्न होती है**, वह सूचना कैसे खोजता है, और संगठन इस प्रक्रिया में कैसे सहायक बनता है।
वैचारिक मॉडल सूचना संगठन को **उपयोगकर्ता-उन्मुख (user-centred)** बनाने का सैद्धांतिक आधार प्रदान करते हैं।

(A) सामान्य वैचारिक मॉडल (General Conceptual Model)

इस मॉडल के अनुसार सूचना खोज की प्रक्रिया निम्न चरणों में होती है:

ज्ञान में अंतर (Knowledge Gap)

⇩

सूचना आवश्यकता की अनुभूति (Information Need)

⇩

सूचना खोज व्यवहार (Information Seeking Behaviour)

⇩

सूचना संगठन एवं प्रणालियाँ (Information Organization & Systems)

⇩

सूचना पुनर्प्राप्ति (Information Retrieval)

⇩

सूचना उपयोग (Information Use)

⇩

संतुष्टि / असंतुष्टि (Satisfaction / Dissatisfaction)

यह मॉडल दर्शाता है कि **यदि सूचना सुव्यवस्थित न हो**, तो यह श्रृंखला टूट जाती है और उपयोगकर्ता की आवश्यकता पूरी नहीं होती।

(B) Wilson का सूचना व्यवहार मॉडल (Wilson's Model of Information Behaviour)

Wilson के अनुसार सूचना आवश्यकता केवल व्यक्तिगत नहीं होती, बल्कि यह—

- मनोवैज्ञानिक कारकों

- सामाजिक परिवेश
- कार्य-परिस्थिति
- पर्यावरणीय बाधाओं

से प्रभावित होती है।

इस मॉडल में सूचना संगठन को एक **सहायक तंत्र (facilitating system)** माना गया है, जो बाधाओं को कम कर सूचना तक पहुँच आसान बनाता है।

(C) Taylor का Information Use Environment (IUE) मॉडल:

Taylor के अनुसार सूचना आवश्यकता को **कार्य-परिस्थिति (Work Context)** से अलग नहीं देखा जा सकता।
इस मॉडल में चार प्रमुख घटक होते हैं:

1. उपयोगकर्ता (User)
2. समस्या या कार्य (Problem / Task)
3. कार्य-परिस्थिति (Environment)
4. सूचना प्रणाली (Information System)

यह मॉडल स्पष्ट करता है कि **सूचना संगठन तभी प्रभावी होता है जब वह उपयोग-परिस्थिति के अनुरूप हो**।

(D) संज्ञानात्मक (Cognitive) मॉडल

इस मॉडल के अनुसार सूचना आवश्यकता उपयोगकर्ता के मन में उत्पन्न **ज्ञान-असंतुलन (Cognitive Uncertainty)** का परिणाम होती है।
सूचना संगठन इस असंतुलन को कम करने में सहायक होता है।

1.11 सूचना संसाधनों/स्रोतों का चयन: मानदंड (Selection Criteria): सभी सूचना स्रोत समान नहीं होते; चयन कुछ मानदंडों पर आधारित होता है। सभी सूचना स्रोत समान गुणवत्ता के नहीं होते। इसलिए सूचना संसाधनों का चयन करते समय प्रामाणिकता, सटीकता, अद्यतनता और प्रासंगिकता जैसे मानदंडों पर विचार करना आवश्यक होता है। सही चयन सूचना संगठन की प्रभावशीलता को बढ़ाता है।

प्रामाणिकता (Authority)

- सटीकता (Accuracy)

- अद्यतनता (Currency)
- प्रासंगिकता (Relevance)
- उपयोगिता (Utility)

1.12 सूचना अधिभार एवं संगठन की उपयोगिता (Information Overload):

सूचना की अधिकता निर्णय को कठिन बना देती है—इसे **Information Overload** कहते हैं। सूचना अधिभार वह स्थिति है जब

"सूचना की अधिकता निर्णय को कठिन बना देती है।"
बौद्धिक संगठन इस समस्या को **फ़िल्टरिंग, वर्गीकरण और अनुक्रमण** द्वारा नियंत्रित करता है।

1.13 सूचना स्रोत बनाम सूचना संसाधन (Sources vs Resources):

दोनों शब्द प्रायः एक-दूसरे के स्थान पर उपयोग होते हैं, परंतु अर्थ में अंतर है।

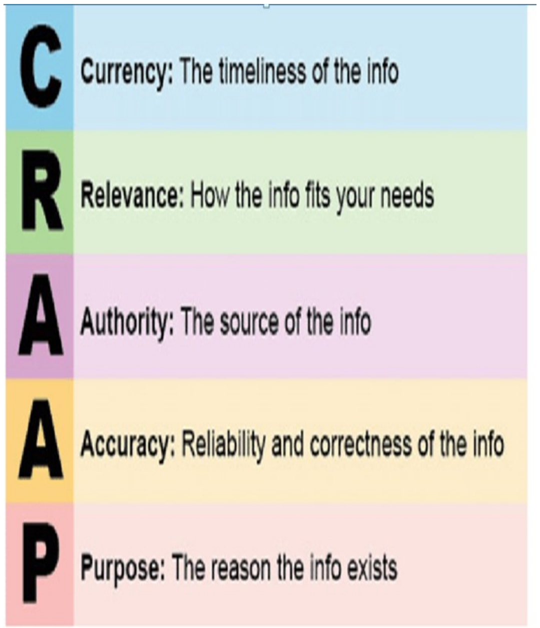
- **सूचना स्रोत**: जहाँ से सूचना प्राप्त होती है
- **सूचना संसाधन**: वह संगठित साधन जिससे सूचना का उपयोग किया जाता है

1.14 सूचना स्रोतों की विशेषताएँ (Characteristics of Information Sources):

अच्छे सूचना स्रोत की पहचान उसकी गुणवत्ता से होती है। प्रामाणिकता, सटीकता, अद्यतनता और उपयोगिता जैसे गुण यह निर्धारित करते हैं कि कोई स्रोत अध्ययन और शोध के लिए कितना उपयुक्त है।

बिंदुवार विस्तार:

- Authority
- Accuracy
- Currency
- Relevance
- Accessibility



1.15 सारांश

इस इकाई में **सूचना के बौद्धिक संगठन** की अवधारणा, विकास, महत्व तथा उसके व्यावहारिक उपयोग का अध्ययन प्रस्तुत किया गया है। यह स्पष्ट होता है कि सूचना का संगठन केवल पुस्तकालयों तक सीमित न होकर सूचना को **खोजने योग्य, समझने योग्य और उपयोगी** बनाने की एक आवश्यक बौद्धिक प्रक्रिया है।

वर्तमान सूचना युग में सुव्यवस्थित सूचना के बिना प्रभावी **सूचना पुनर्प्राप्ति** और **ज्ञान निर्माण** संभव नहीं है। इसलिए सूचना का बौद्धिक संगठन न केवल पुस्तकालयों, बल्कि संपूर्ण **ज्ञान समाज** के लिए अत्यंत आवश्यक है। सुव्यवस्थित सूचना ही प्रभावी ज्ञान के विकास की आधारशिला बनती है।

1.16 शब्दावली (Glossary):

1. **सूचना (Information):** डेटा का वह अर्थपूर्ण रूप जो संदर्भ और उपयोगिता से युक्त होकर उपयोगकर्ता की आवश्यकता को पूरा करता है।
2. **डेटा (Data)**: कच्चे तथ्य या आँकड़े, जिन्हें संसाधित करने पर सूचना प्राप्त होती है।
3. **बौद्धिक संगठन (Intellectual Organization):** सूचना के विषय, अवधारणा और ज्ञानात्मक संरचना के आधार पर किया गया तार्किक एवं विश्लेषणात्मक संगठन।

4. **सूचना संगठन (Information Organization):** सूचना को वर्गीकरण, अनुक्रमण और वर्णन के माध्यम से खोजने योग्य बनाने की प्रक्रिया।
5. **ज्ञान संगठन (Knowledge Organization):** ज्ञान की अवधारणाओं और उनके आपसी संबंधों को संरचित करने की प्रक्रिया।
6. **विषय विश्लेषण (Subject Analysis):** किसी दस्तावेज़ के मुख्य विषय एवं उपविषयों की पहचान की प्रक्रिया।
7. **वर्गीकरण (Classification):** समान विषयों वाले संसाधनों को एक निश्चित योजना के अंतर्गत व्यवस्थित करना।
8. **अनुक्रमण (Indexing):**विषय शब्दों/की-वर्ड के माध्यम से सूचना तक शीघ्र पहुँच सुनिश्चित करने की तकनीक।
9. **सूचना आवश्यकता (Information Need):** उपयोगकर्ता के वर्तमान ज्ञान और अपेक्षित ज्ञान के बीच का अंतर।
10. **सूचना अधिभार (Information Overload):** सूचना की अत्यधिक मात्रा से उत्पन्न वह स्थिति जिसमें निर्णय लेना कठिन हो जाता है।
11. **प्रासंगिकता (Relevance):** सूचना का उपयोगकर्ता की आवश्यकता से मेल खाना।
12. **सूचना स्रोत (Information Source):** वह माध्यम जहाँ से सूचना प्राप्त होती है, जैसे पुस्तक, पत्रिका, वेबसाइट।
13. **सूचना संसाधन (Information Resource):** संगठित एवं संरचित सूचना सामग्री जिसका उपयोग अध्ययन व शोध में होता है।
14. **उपयोगकर्ता-केंद्रित संगठन (User-centred Organization):** ऐसा संगठन जिसमें उपयोगकर्ता की आवश्यकता और व्यवहार को प्राथमिकता दी जाती है।
15. **सूचना पुनर्प्राप्ति (Information Retrieval):** संगठित सूचना संग्रह से आवश्यक सूचना को खोजकर प्राप्त करने की प्रक्रिया।

10.17 निबंधात्मक प्रश्न (Essay Type Questions) – उत्तर सहित

प्रश्न 1. सूचना का बौद्धिक संगठन क्या है? इसके महत्व की विवेचना कीजिए।

उत्तर : सूचना का बौद्धिक संगठन वह प्रक्रिया है जिसमें सूचना को उसके विषय, अर्थ और संदर्भ के आधार पर व्यवस्थित किया जाता है। इसका महत्व इस बात में है कि यह सूचना को खोजने योग्य बनाता है, शोध की गुणवत्ता बढ़ाता है और सूचना अधिभार को नियंत्रित करता है।

प्रश्न 2. आधुनिक सूचना समाज में सूचना संगठन की भूमिका स्पष्ट कीजिए।

उत्तर:
सूचना समाज में सूचना की मात्रा अत्यधिक है। सूचना संगठन उपयोगकर्ता को सही समय पर सही सूचना प्रदान कर ज्ञान निर्माण, निर्णय-निर्माण और शोध को प्रभावी बनाता है।

प्रश्न 3. सूचना अधिभार क्या है? बौद्धिक संगठन इसके समाधान में कैसे सहायक है?

उत्तर:
सूचना अधिभार वह स्थिति है जब सूचना की अधिकता निर्णय को कठिन बना देती है। बौद्धिक संगठन वर्गीकरण, फ़िल्टरिंग और अनुक्रमण द्वारा इस समस्या को कम करता है।

प्रश्न 4. सूचना आवश्यकता और सूचना संगठन के बीच संबंध स्पष्ट कीजिए।

उत्तर:
सूचना आवश्यकता ज्ञान में विद्यमान रिक्तता से उत्पन्न होती है। सूचना संगठन इस रिक्तता को भरने का माध्यम बनता है।

प्रश्न 5. डिजिटल पुस्तकालयों में सूचना संगठन का महत्व लिखिए।

उत्तर:
डिजिटल पुस्तकालयों में प्रभावी खोज, तेज़ पहुँच और वैश्विक संसाधन-साझाकरण सूचना संगठन के बिना संभव नहीं है।

10.19 लघु प्रश्न (Short Questions) – उत्तर सहित

1. **सूचना संगठन से आप क्या समझते हैं?**
   → सूचना को खोजने योग्य बनाने की बौद्धिक प्रक्रिया।
2. **बौद्धिक संगठन क्यों आवश्यक है?**
   → सूचना अधिभार कम करने और खोज को सरल बनाने के लिए।
3. **सूचना स्रोत और संसाधन में एक अंतर लिखिए।**
   → स्रोत सूचना का माध्यम है, संसाधन संगठित रूप।
4. **प्रासंगिकता का अर्थ लिखिए।**
   → सूचना का आवश्यकता से मेल खाना।
5. **सूचना पुनर्प्राप्ति क्या है?**
   → आवश्यक सूचना को खोजकर प्राप्त करने की प्रक्रिया।

MCQ– उत्तर सहित

1. सूचना संगठन का मुख्य उद्देश्य है—
   a) संग्रह बढ़ाना
   b) सूचना को खोजने योग्य बनाना ☑
   c) सूचना छिपाना
   d) केवल संरक्षण
2. बौद्धिक संगठन मुख्यतः किस प्रक्रिया से जुड़ा है?
   a) भौतिक
   b) तकनीकी
   c) मानसिक/बौद्धिक ☑
   d) प्रशासनिक
3. सूचना अधिभार का संबंध है—
   a) सूचना की कमी
   b) सूचना की अधिकता ☑
   c) सूचना सुरक्षा
   d) सूचना संरक्षण
4. अनुक्रमण का उद्देश्य है—
   a) संरक्षण
   b) तेज़ खोज ☑
   c) मुद्रण
   d) संग्रह
5. सूचना आवश्यकता उत्पन्न होती है—
   a) संतुष्टि से
   b) ज्ञान अंतर से ☑
   c) संग्रह से
   d) वर्गीकरण से
6. Knowledge Organization का संबंध है—
   a) डेटा से
   b) अवधारणाओं से ☑
   c) हार्डवेयर से
   d) नेटवर्क से
7. उपयोगकर्ता-केंद्रित संगठन में प्राथमिकता होती है—
   a) संग्रह
   b) तकनीक

c) उपयोगकर्ता ☑
d) लागत

8. सूचना पुनर्प्राप्ति का आधार है—
a) वर्गीकरण और अनुक्रमण ☑
b) संरक्षण
c) मुद्रण
d) प्रचार

9. प्रासंगिकता का संबंध है—
a) मात्रा से
b) आवश्यकता से ☑
c) लागत से
d) भाषा से

10. डिजिटल युग में सूचना संगठन है—
a) अनावश्यक
b) सीमित
c) अत्यंत आवश्यक ☑
d) अप्रासंगिक

10.20 संदर्भ ग्रंथ सूची (References)

1. Chowdhury, G. G. Introduction to Modern Information Retrieval.
2. Lancaster, F. W. Indexing and Abstracting in Theory and Practice.
3. Hjørland, B. Knowledge Organization.
4. Rowley, J. Organizing Knowledge.
5. Bawden, D. & Robinson, L. Introduction to Information Science.
6. Taylor, R. S. Information Use Environments.
7. Vickery, B. C. Classification and Indexing in Science.
8. Svenonius, E. The Intellectual Foundation of Information Organization.
9. Bates, M. J. Information and the Information Professions.
10. Ingwersen, P. Information Retrieval Interaction.
11. रंगनाथन, एस. आर. पुस्तकालय वर्गीकरण.
12. त्रिपाठी, एस. एन. पुस्तकालय एवं सूचना विज्ञान के मूल सिद्धांत.
13. शर्मा, जे. पी. सूचना विज्ञान : सिद्धांत एवं व्यवहार.
14. मिश्रा, आर. एन. पुस्तकालय एवं सूचना सेवाएँ.
15. पाण्डेय, बी. एन. सूचना संगठन एवं पुनर्प्राप्ति.
16. कुमार, पी. एस. जी. सूचना पुनर्प्राप्ति प्रणालियाँ.
17. शुक्ल, के. एन. सूचना स्रोत एवं सेवाएँ.

18. वर्मा, आर. के. पुस्तकालय प्रबंधन.
19. सिंह, एस. पी. आधुनिक पुस्तकालय विज्ञान.
20. उपाध्याय, एम. के. सूचना युग एवं पुस्तकालय.

इकाई–2 : अनुक्रमण भाषाएँ—अवधारणाएँ और प्रकार, विषय शीर्षकसूचियाँ एवं थिसॉरस (Indexing Languages: Concepts, Types, Subject Headings, and Thesaurus)

इकाई की रूपरेखा

2.1 प्रस्तावना
2.2 उद्देश्य
2.3 अनुक्रमण भाषा: अर्थ एवं परिभाषा **(Indexing Language)**
2.4 अनुक्रमण भाषाओं की आवश्यकता एवं भूमिका **(Need & Role)**
2.5 अनुक्रमण भाषाओं के प्रकार **(Types)**
2.5.1 नियंत्रित शब्दावली **(Controlled Vocabulary)**
2.5.2 अनियंत्रित शब्दावली/की-वर्ड **(Uncontrolled / Keywords)**
2.6 शब्द नियंत्रण **(Vocabulary Control):** सिद्धांत एवं उपकरण
2.7 विषय शीर्षकसूचियाँ: अवधारणा, उद्देश्य, विशेषताएँ
2.8 विषय शीर्षक रचना **(Subject Heading Construction)** एवं **Cross References**
2.9 थिसॉरस: परिभाषा, संरचना, शब्द-संबंध **(BT/NT/RT/UF)**
2.10 थिसॉरस निर्माण के चरण **(Thesaurus Construction Steps)**
2.11 थिसॉरस के प्रकार एवं उपयोग क्षेत्र
2.12 अनुक्रमण भाषाओं की सीमाएँ एवं गुणवत्ता मुद्दे
2.13 डिजिटल परिवेश में अनुक्रमण भाषा **(Metadata, Tags, Folksonomy)**
2.14 केस/उदाहरण: विषय शीर्षक बनाम थिसॉरस टर्म
2.15 सारांश
2.16 शब्दावली **(Glossary)**

2.17 निबंधात्मक प्रश्न एवं उत्तर
2.18 लघु प्रश्न **(Short Questions)**
2.19 बहुविकल्पीय प्रश्न **(MCQs)**

2.20 संदर्भ ग्रंथ सूची

इकाई–2 : अनुक्रमण भाषाएँ—अवधारणाएँ और प्रकार, विषय शीर्षकसूचियाँ एवं थिसॉरस (Indexing Languages: Concepts & Types, Subject Heading Lists and Thesaurus)

2.1 प्रस्तावना (Introduction)

सूचना विस्फोट (Information Explosion) के वर्तमान युग में पुस्तकालयों, डिजिटल रिपॉज़िटरी, ऑनलाइन डेटाबेस तथा वेब संसाधनों में सूचनाओं की मात्रा तीव्र गति से बढ़ रही है। ऐसी स्थिति में सूचना का केवल संग्रह पर्याप्त नहीं है; उसे **इस प्रकार संगठित करना आवश्यक है कि उपयोगकर्ता न्यूनतम समय में सटीक और प्रासंगिक सूचना प्राप्त कर सके।** यहीं पर **अनुक्रमण भाषाएँ (Indexing Languages)** अत्यंत महत्वपूर्ण भूमिका निभाती हैं।

अनुक्रमण भाषाएँ ज्ञान संगठन प्रणालियों (Knowledge Organization Systems – KOS) का एक प्रमुख घटक हैं, जो दस्तावेज़ों की विषय-वस्तु को **मानकीकृत शब्दों, पदों और अवधारणाओं** के माध्यम से अभिव्यक्त करती हैं। ये भाषाएँ दस्तावेज़ और उपयोगकर्ता के बीच एक **बौद्धिक सेतु (Intellectual Bridge)** का कार्य करती हैं।

2.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् शिक्षार्थी:

1. अनुक्रमण भाषा की अवधारणा एवं परिभाषा समझ सकेंगे
2. अनुक्रमण भाषाओं की आवश्यकता और भूमिका का विश्लेषण कर सकेंगे
3. Controlled एवं Uncontrolled Vocabulary में अंतर स्पष्ट कर सकेंगे
4. Vocabulary Control के सिद्धांतों (Canons) को गहराई से समझ सकेंगे
5. Subject Heading Lists एवं Thesaurus की संरचना और उपयोग जान सकेंगे
6. डिजिटल परिवेश में अनुक्रमण भाषाओं की भूमिका का मूल्यांकन कर सकेंगे

2.3 अनुक्रमण भाषा: अर्थ एवं परिभाषा (Meaning and Definition of Indexing Language)

अर्थ (Meaning)

अनुक्रमण भाषा वह **कृत्रिम अथवा नियंत्रित भाषा** है, जिसके माध्यम से किसी दस्तावेज़ की विषय-वस्तु को ऐसे शब्दों या पदों में व्यक्त किया जाता है, जो **सूचना खोज प्रणाली (Information Retrieval System)** द्वारा पहचाने जा सकें।

परिभाषाएँ (Definitions)

- **According to Foskett (1971):**

  "An indexing language is a set of terms used to represent the subject content of documents."

- **According to Lancaster:**

  "Indexing language provides a controlled means of translating concepts from documents into searchable terms."

इन परिभाषाओं से स्पष्ट है कि अनुक्रमण भाषा **विषय-अभिव्यक्ति का मानकीकृत माध्यम** है।

2.4 अनुक्रमण भाषाओं की आवश्यकता एवं भूमिका (Need & Role of Indexing Languages)

(क) आवश्यकता (Need)

1. समानार्थक शब्दों (Synonyms) की समस्या
2. बहुअर्थी शब्दों (Homonyms) से उत्पन्न भ्रम
3. विषय की सटीक और एकरूप अभिव्यक्ति
4. Recall एवं Precision में सुधार
5. बड़े डेटाबेस में प्रभावी खोज

(ख) भूमिका (Role)

- दस्तावेज़ों को विषयानुसार जोड़ना
- OPAC एवं डेटाबेस में सटीक खोज
- उपयोगकर्ता की सूचना आवश्यकताओं को पूरा करना

उदाहरण:
यदि उपयोगकर्ता "Artificial Intelligence" खोजता है और दस्तावेज़ में "AI" लिखा है, तो Controlled Indexing Language दोनों को जोड़ देती है।

2.5 अनुक्रमण भाषाओं के प्रकार (Types of Indexing Languages)

अनुक्रमण भाषाएँ मुख्यतः दो प्रकार की होती हैं:

2.5.1 नियंत्रित शब्दावली (Controlled Vocabulary)

परिभाषा:
Controlled Vocabulary वह शब्दावली है जिसमें **पूर्व-निर्धारित, मानकीकृत और स्वीकृत शब्दों** का ही प्रयोग किया जाता है।

मुख्य विशेषताएँ:

- Authority Control
- Consistency
- Synonym Control
- बेहतर Precision

उदाहरण:

- Library of Congress Subject Headings (LCSH)
- Medical Subject Headings (MeSH)

2.5.2 अनियंत्रित शब्दावली / की-वर्ड (Uncontrolled Vocabulary / Keywords)

परिभाषा:
जहाँ अनुक्रमक या उपयोगकर्ता प्राकृतिक भाषा के किसी भी शब्द का प्रयोग कर सकता है।

विशेषताएँ:

- User-friendly
- कोई मानकीकरण नहीं
- Inconsistency की समस्या

उदाहरण:
Web search keywords, Social media tags

2.6 शब्द नियंत्रण (Vocabulary Control): सिद्धांत एवं उपकरण

(Canons of Vocabulary Control – Elaborated)

Vocabulary Control का अर्थ

Vocabulary Control का तात्पर्य है—

विषय अभिव्यक्ति के लिए शब्दों का **नियंत्रित, सुसंगत और उपयोगकर्ता-उन्मुख चयन**।

मुख्य सिद्धांत (Canons)

1. Canon of Consistency: एक ही विषय के लिए एक ही Preferred Term का प्रयोग।

2. Canon of Relevance: चुना गया शब्द विषय से प्रत्यक्ष रूप से संबंधित हो।

3. Canon of Ascertainability: शब्द ऐसा हो जिसे उपयोगकर्ता आसानी से समझ सके।

4. Canon of Hospitality: नई अवधारणाओं और विषयों को समाहित करने की क्षमता।

5. Canon of User Orientation: उपयोगकर्ता की खोज-भाषा को ध्यान में रखना।

2.7 विषय शीर्षकसूचियाँ (Subject Heading Lists)

अवधारणा

Subject Heading List नियंत्रित शब्दों की वह सूची है, जो **विषय-अभिव्यक्ति के लिए मानक शीर्षक** प्रदान करती है।

उद्देश्य

- विषय की एकरूपता
- Catalog/OPAC में सटीक खोज

विशेषताएँ

- Pre-coordinated
- Authority control
- Cross references

उदाहरण:
Library of Congress Subject Headings (LCSH)

2.8 विषय शीर्षक रचना एवं Cross References

Subject Heading Construction

मुख्य विषय + उपविषय + स्थान/काल

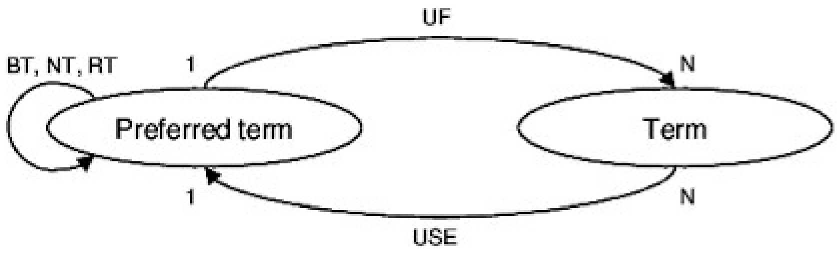
उदाहरण:
Libraries — India — Automation

Cross References

- **See:** पर्यायवाची
- **See also:** संबंधित शब्द

2.9 थिसॉरस: परिभाषा, संरचना एवं शब्द-संबंध (Thesaurus: Definition, Structure & Relationships)

थिसॉरस एक **नियंत्रित और संरचित शब्दावली उपकरण** है, जिसमें शब्दों को उनके **अर्थगत संबंधों** के आधार पर व्यवस्थित किया जाता है।



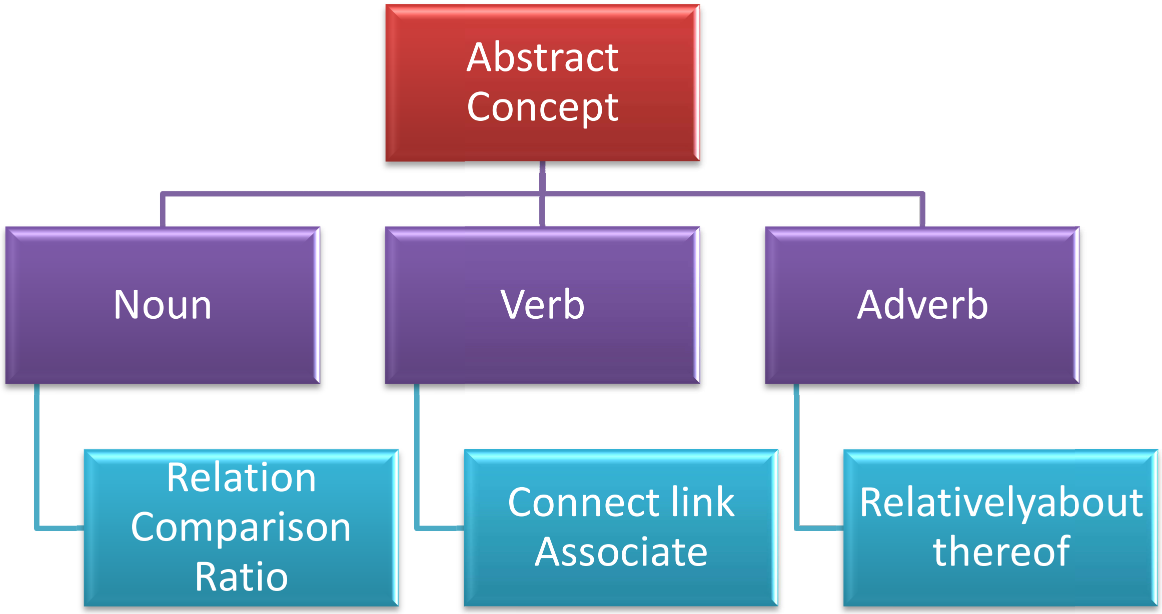
शब्द-संबंध:

- BT (Broader Term)
- NT (Narrower Term)
- RT (Related Term)
- UF (Used For)

Concept-based Thesaurus Representation (Text Diagram)

Example: CONCEPT: Climate Change (PT)

- UF : Global Warming
- BT : Environmental Change
- NT : Greenhouse Effect
- NT : Sea Level Rise
- RT : Sustainable Development



ऊपर "RELATION (Abstract Concept)" लिखें, नीचे तीन शाखाएँ (Noun/Verb/Adverb) और 2–3 उदाहरण पर्याप्त हैं।

- **Concept-based representation in Thesaurus (Abstract Concept vs Word Forms)"**

नीचे दिया गया आरेख दर्शाता है कि Thesaurus में एक ही abstract concept विभिन्न grammatical forms में व्यक्त शब्दों से कैसे जुड़ा होता है। यह संरचना semantic control सुनिश्चित करती है और सूचना पुनःप्राप्ति की गुणवत्ता को बढ़ाती है।

2.10 थिसॉरस निर्माण के चरण

1. विषय क्षेत्र का चयन
2. शब्द संग्रह
3. शब्द विश्लेषण
4. संबंध निर्धारण
5. मानकीकरण
6. परीक्षण एवं संशोधन

2.11 थिसॉरस के प्रकार एवं उपयोग क्षेत्र

प्रकार

- General Thesaurus
- Subject-specific Thesaurus
- Multilingual Thesaurus

उपयोग

- Digital libraries
- Online databases
- Information Retrieval Systems

2.12 अनुक्रमण भाषाओं की सीमाएँ एवं गुणवत्ता मुद्दे

सीमाएँ:

- अद्यतन की समस्या
- बहुविषयी विषय
- भाषा एवं सांस्कृतिक पक्षपात

गुणवत्ता मापदंड:

- Consistency
- Specificity
- Exhaustivity

2.13 डिजिटल परिवेश में अनुक्रमण भाषा (Indexing Language in Digital Environment):

डिजिटल सूचना परिवेश में अनुक्रमण भाषाएँ पारंपरिक नियंत्रित शब्दावली तक सीमित न रहकर **मेटाडाटा, टैगिंग और उपयोगकर्ता-आधारित शब्दावली** के रूप में विकसित हो गई हैं। इनका मुख्य उद्देश्य ऑनलाइन प्रणालियों में **तेज़, सटीक और लचीली सूचना पुनःप्राप्ति** सुनिश्चित करना है।

(i) Metadata – Dublin Core

Metadata को "डेटा का डेटा" (Data About Data) कहा जाता है।

Dublin Core एक मानक मेटाडाटा स्कीमा है, जिसमें *Title, Creator, Subject, Description* जैसे 15 मूल तत्व होते हैं।

- यह डिजिटल दस्तावेज़ों की **विषय-पहचान**, **इंटरऑपरेबिलिटी** और **सटीक खोज** में सहायता करता है।
- डिजिटल पुस्तकालयों, रिपॉज़िटरी और OPAC में इसका व्यापक उपयोग होता है।

(ii) Tags – User-generated

- Tags वे की-वर्ड होते हैं जिन्हें **उपयोगकर्ता स्वयं** दस्तावेज़ या वेब सामग्री से जोड़ते हैं।
- ये **natural language आधारित**, सरल और user-friendly होते हैं। सोशल मीडिया, ब्लॉग, और वेब प्लेटफॉर्म पर खोज को सहज बनाते हैं।

हालाँकि, टैगिंग में **मानकीकरण का अभाव** होने के कारण समान विषय के लिए अलग-अलग शब्द प्रयोग की समस्या उत्पन्न होती है।

(iii) Folksonomy – लोकतांत्रिक पर असंगत

Folksonomy वह प्रणाली है जिसमें टैगिंग पूरी तरह **उपयोगकर्ताओं द्वारा स्वतंत्र रूप से** की जाती है।

- इसे **लोकतांत्रिक (democratic)** इसलिए कहा जाता है क्योंकि इसमें किसी केंद्रीय नियंत्रण या विशेषज्ञ शब्दावली की आवश्यकता नहीं होती।
- परंतु यह **असंगत (inconsistent)** होती है, क्योंकि अलग-अलग उपयोगकर्ता एक ही विषय के लिए अलग शब्दों का प्रयोग कर सकते हैं।

- **निष्कर्षतः**, Folksonomy खोज को व्यापक बनाती है (High Recall), परंतु सटीकता (Precision) नियंत्रित शब्दावली की तुलना में कम हो सकती है।

2.14 केस / उदाहरण : विषय शीर्षक बनाम थिसॉरस टर्म (Case Studies: Subject Headings vs Thesaurus Terms)

2.14.1 भूमिका (Introduction to the Case)

अनुक्रमण भाषाओं के अध्ययन में **Subject Heading Lists** और **Thesaurus** दोनों महत्वपूर्ण उपकरण हैं, किंतु उनकी **संरचना, समन्वय प्रक्रिया और उपयोग क्षेत्र** में स्पष्ट अंतर पाया जाता है।
जहाँ Subject Headings पूर्व-समन्वित (pre-coordinated) रूप में विषय अभिव्यक्ति प्रदान करते हैं, वहीं Thesaurus अवधारणा-आधारित (concept-based) और अर्थगत संबंधों पर आधारित होता है।

इस अनुभाग में **व्यावहारिक केस/उदाहरणों** के माध्यम से यह स्पष्ट किया गया है कि **एक ही विषय** को Subject Heading और Thesaurus में किस प्रकार भिन्न-भिन्न रूपों में अभिव्यक्त किया जाता है।
इसका उद्देश्य शिक्षार्थियों को यह समझने में सहायता देना है कि **किस संदर्भ—catalog, OPAC या digital information system—में कौन-सा अनुक्रमण उपकरण अधिक उपयुक्त सिद्ध होता है।**

2.14.2 केस/उदाहरण: विषय शीर्षक बनाम थिसॉरस टर्म

आधार	Subject Heading	Thesaurus
संरचना	Pre-coordinated	Post-coordinated
नियंत्रण	सीमित	अधिक नियंत्रित
उदाहरण	LCSH	MeSH

उदाहरण विषय: *Digital Libraries*

- **Subject Heading:** Libraries—Digital
- **Thesaurus:** Digital Libraries (RT: Information Systems)

2.14.3 केस–2 : "Climate Change" विषय का विश्लेषण

(A) Subject Heading के रूप में

LCSH शैली में Subject Heading:
Climate change — Environmental aspects — India
विशेषताएँ:

- Pre-coordinated heading
- शब्दों का क्रम पहले से तय
- Printed catalogues और OPAC के लिए उपयुक्त

सीमा:
यदि उपयोगकर्ता केवल "Global Warming" खोजे, तो परिणाम छूट सकते हैं।

(B) Thesaurus Term के रूप में

Thesaurus representation:
- **Preferred Term (PT):** Climate Change
- **UF:** Global Warming

- **BT:** Environmental Change
- **NT:** Greenhouse Effect, Sea Level Rise
- **RT:** Sustainable Development

विशेषताएँ:

- Concept-based linking
- Synonym control
- Digital databases में उच्च Recall

2.14.3 केस–3 : "Artificial Intelligence" विषय

Subject Heading

Artificial intelligence — Applications — Education

Thesaurus Terms

- PT: Artificial Intelligence
- UF: AI
- NT: Machine Learning, Expert Systems
- RT: Data Science, Robotics

निष्कर्ष:

Thesaurus **semantic search** को बेहतर बनाता है।

2.14.4 तुलनात्मक विश्लेषण (Comparative Analysis)

संरचना	Pre-coordinated	Post-coordinated
लचीलापन	सीमित	अधिक
डिजिटल उपयुक्तता	मध्यम	उच्च
Recall	अपेक्षाकृत कम	अधिक
Precision	अच्छा	अच्छा (यदि सही उपयोग)

2.14.5 सरल अवधारणात्मक आरेख (Conceptual Diagram)

Subject Heading vs Thesaurus (Clean Text Diagram)

Subject Heading (Pre-coordinated)

Artificial intelligence — Education — India

Thesaurus (Post-coordinated)

PT : Artificial Intelligence
UF : AI
BT : Computer Science
NT : Machine Learning
RT : Educational Technology

व्याख्या:
यह आरेख स्पष्ट करता है कि **Subject Heading** एक निश्चित एवं पूर्व-समन्वित शीर्षक प्रदान करता है, जबकि **Thesaurus** एक ही अवधारणा को उससे संबंधित व्यापक, संकीर्ण एवं समानार्थक शब्दों से जोड़कर **अधिक लचीली और अर्थ-आधारित विषय अभिव्यक्ति** प्रस्तुत करता है।

2.4.6 परीक्षा दृष्टि से विश्लेषणात्मक टिप्पणी

आधुनिक डिजिटल सूचना प्रणालियों में Thesaurus, Subject Heading की तुलना में अधिक प्रभावी सिद्ध होता है, क्योंकि यह उपयोगकर्ता की खोज-भाषा को बेहतर ढंग से समायोजित करता है और semantic relationships को प्रदर्शित करता है।

2.15 सारांश (Summary)

यह इकाई अनुक्रमण भाषाओं की अवधारणा, प्रकार, सिद्धांत, **Subject Heading Lists** तथा **Thesaurus** का समग्र और विश्लेषणात्मक अध्ययन प्रस्तुत करती है। अध्ययन से यह स्पष्ट होता है कि आधुनिक सूचना तंत्र में **Controlled Vocabulary** विषय-अभिव्यक्ति में एकरूपता प्रदान करती है, जबकि **Thesaurus** अर्थगत संबंधों के माध्यम से खोज की प्रभावशीलता को बढ़ाता है। इस प्रकार, दोनों उपकरण मिलकर प्रभावी **सूचना संगठन एवं पुनःप्राप्ति** की आधारशिला निर्मित करते हैं।

2.16 शब्दावली (Glossary)

- **Indexing Language (अनुक्रमण भाषा)**: दस्तावेज़ की विषय-वस्तु को मानकीकृत शब्दों में व्यक्त करने की प्रणाली।
- **Controlled Vocabulary (नियंत्रित शब्दावली)**: पूर्व-निर्धारित और स्वीकृत शब्दों का संग्रह, जो विषय-अभिव्यक्ति में एकरूपता लाता है।
- **Thesaurus (थिसॉरस)**: अवधारणाओं को उनके अर्थगत संबंधों (BT/NT/RT/UF) के आधार पर व्यवस्थित करने वाला नियंत्रित शब्दकोश।

- **Metadata (मेटाडाटा)**: वह संरचित जानकारी जो किसी सूचना संसाधन का वर्णन करती है (Data about Data)।
- **Folksonomy (फोक्सोनॉमी)**: उपयोगकर्ताओं द्वारा स्वतंत्र रूप से की गई टैगिंग पर आधारित अनुक्रमण प्रणाली।
- **UF (Used For)**: Thesaurus में वह शब्द जो **Preferred Term** के स्थान पर प्रयुक्त नहीं होता, बल्कि उसी अवधारणा की ओर संकेत करता है।
- **Preferred Term (PT)**: Thesaurus में किसी अवधारणा को अभिव्यक्त करने के लिए **स्वीकृत एवं मानक शब्द**।
- **Broader Term (BT):** वह शब्द जो किसी अवधारणा से **अधिक व्यापक** श्रेणी को दर्शाता है।
- **Narrower Terma (NT):** वह शब्द जो किसी अवधारणा से **अधिक विशिष्ट या संकीर्ण** अर्थ को दर्शाता है।
- **Related Term (RT):** वह शब्द जो किसी अवधारणा से **अर्थगत रूप से संबंधित** हो, लेकिन न तो व्यापक हो न संकीर्ण।
- **Pre-coordination:** अनुक्रमण की वह प्रक्रिया जिसमें विषय को **पहले से निश्चित क्रम** में एक ही शीर्षक के रूप में व्यक्त किया जाता है।
- **Post-coordination:** वह प्रक्रिया जिसमें अलग-अलग शब्दों को **खोज के समय** संयोजित किया जाता है (Thesaurus आधारित प्रणाली)।

Semantic Relationships

→ शब्दों के बीच अर्थगत संबंध, जैसे BT/NT/RT/UF, जो Thesaurus की मूल विशेषता हैं।

Post-coordination

→ वह प्रक्रिया जिसमें अलग-अलग शब्दों को **खोज के समय** संयोजित किया जाता है (Thesaurus आधारित प्रणाली)।

Semantic Relationships

→ शब्दों के बीच अर्थगत संबंध, जैसे BT/NT/RT/UF, जो Thesaurus की मूल विशेषता हैं।

Query Expansion

- **Semantic Relationships:** शब्दों के बीच अर्थगत संबंध, जैसे BT/NT/RT/UF, जो Thesaurus की मूल विशेषता हैं।
- **Query Expansion:** खोज प्रक्रिया में समानार्थक, व्यापक या संबंधित शब्दों को जोड़कर **Recall बढ़ाने की तकनीक**।
- **Recall:** खोज में प्राप्त कुल प्रासंगिक दस्तावेज़ों का अनुपात।
- **Precision:** खोज में प्राप्त दस्तावेज़ों में से **वास्तव में प्रासंगिक दस्तावेज़ों का** अनुपात।
- **Cross References (See / See also):** Subject Heading Lists में प्रयुक्त संदर्भ, जो समानार्थक या संबंधित विषयों की ओर मार्गदर्शन करते हैं।
- **ISO 25964:** Thesaurus निर्माण और विभिन्न vocabularies के interoperability हेतु अंतरराष्ट्रीय मानक।
- **Ambiguity:** एक ही शब्द के **एक से अधिक अर्थ** होने की स्थिति, जिसे Thesaurus कम करने में सहायक होता है।
- **Consistency in Indexing:** समान विषय के लिए **समान शब्दों के प्रयोग** की स्थिति, जो Controlled Vocabulary का प्रमुख उद्देश्य है।

2.17 निबंधात्मक प्रश्न (Essay Questions & Ans)

प्रश्न 1. विषय शीर्षक और थिसॉरस में अंतर को उदाहरण सहित स्पष्ट कीजिए।

उत्तर: Subject Heading और Thesaurus दोनों नियंत्रित शब्दावली उपकरण हैं, किंतु उनकी संरचना और कार्य-प्रणाली भिन्न होती है।
Subject Heading पूर्व-समन्वित (pre-coordinated) होता है, जहाँ पूरा विषय एक निश्चित शीर्षक के रूप में व्यक्त किया जाता है।
इसके विपरीत, **Thesaurus** अवधारणा-आधारित (concept-based) होता है और शब्दों को BT, NT, RT, UF जैसे अर्थगत संबंधों से जोड़ता है।

उदाहरण:

- Subject Heading: *Artificial intelligence — Education*
- Thesaurus:
    - PT: Artificial Intelligence
    - UF: AI
    - NT: Machine Learning
    - RT: Educational Technology

इस प्रकार, Thesaurus डिजिटल खोज में अधिक लचीलापन प्रदान करता है।

प्रश्न 2. डिजिटल पुस्तकालयों में थिसॉरस की भूमिका का मूल्यांकन कीजिए।

उत्तर: डिजिटल पुस्तकालयों में थिसॉरस semantic search को सशक्त बनाता है। यह समानार्थक शब्दों, व्यापक एवं संकीर्ण अवधारणाओं को जोड़कर खोज की गुणवत्ता सुधारता है।
Thesaurus query expansion की सुविधा देता है, जिससे Recall बढ़ता है और उपयोगकर्ता को अधिक प्रासंगिक परिणाम प्राप्त होते हैं।
इस कारण आधुनिक डिजिटल सूचना प्रणालियों में थिसॉरस एक महत्वपूर्ण अनुक्रमण उपकरण है।

प्रश्न 3. क्यों कहा जाता है कि थिसॉरस concept-based indexing को बढ़ावा देता है?

उत्तर: Thesaurus शब्दों की बजाय **अवधारणाओं (concepts)** पर आधारित होता है। यह एक abstract concept को उसके विभिन्न शब्द-रूपों और संबंधित अवधारणाओं से जोड़ता है।
UF, BT, NT और RT संबंधों के माध्यम से यह सुनिश्चित करता है कि उपयोगकर्ता द्वारा प्रयुक्त शब्द चाहे जो हो, वह उसी अवधारणा से जुड़ जाए।
इसी कारण थिसॉरस को concept-based indexing का प्रभावी उपकरण माना जाता है।

प्रश्न 4. Subject Heading Lists की सीमाएँ क्या हैं?

उत्तर: Subject Heading Lists की प्रमुख सीमाएँ निम्नलिखित हैं—

- ये pre-coordinated होती हैं, जिससे लचीलापन कम हो जाता है।
- उपयोगकर्ता की खोज-भाषा से मेल न खाने पर परिणाम छूट सकते हैं।
- डिजिटल परिवेश में semantic relationships का अभाव रहता है।

इन सीमाओं के कारण Thesaurus की आवश्यकता बढ़ जाती है।

प्रश्न 5. एक विषय को Subject Heading और Thesaurus दोनों में कैसे व्यक्त किया जाता है?

उत्तर: एक ही विषय को दोनों उपकरणों में अलग ढंग से व्यक्त किया जाता है।

उदाहरण: **"Climate Change"**

- Subject Heading: *Climate change — Environmental aspects*
- Thesaurus:
    - PT: Climate Change
    - UF: Global Warming
    - BT: Environmental Change
    - NT: Greenhouse Effect

इससे स्पष्ट होता है कि Thesaurus विषय की अधिक व्यापक अभिव्यक्ति करता है।

6. Controlled Vocabulary और Uncontrolled Vocabulary की तुलनात्मक विवेचना कीजिए।

उत्तर (संक्षेप में): Controlled Vocabulary में मानकीकृत और स्वीकृत शब्दों का प्रयोग होता है, जिससे विषय-अभिव्यक्ति में एकरूपता आती है और Precision बढ़ती है। Uncontrolled Vocabulary प्राकृतिक भाषा पर आधारित होती है, जो user-friendly तो होती है, परंतु असंगति और ambiguity की समस्या उत्पन्न करती है। डिजिटल सूचना प्रणालियों में Controlled Vocabulary अधिक विश्वसनीय मानी जाती है।

7. Vocabulary Control के सिद्धांतों (Canons) का प्रभावी अनुक्रमण में क्या महत्व है?

उत्तर: Vocabulary Control के सिद्धांत जैसे Consistency, Relevance और User Orientation यह सुनिश्चित करते हैं कि विषयों की अभिव्यक्ति स्पष्ट, सुसंगत और उपयोगकर्ता-उन्मुख हो।
इन सिद्धांतों के अभाव में अनुक्रमण में असमानता और retrieval failure की संभावना बढ़ जाती है।

8. Subject Heading Lists को एक Knowledge Organization Tool के रूप में समझाइए।

उत्तर: Subject Heading Lists नियंत्रित शब्दावली पर आधारित उपकरण हैं, जो विषयों को पूर्व-समन्वित शीर्षकों के रूप में प्रस्तुत करते हैं।
ये catalog और OPAC में एकरूप विषय-प्रवेश (uniform access) सुनिश्चित करती हैं, जिससे सूचना खोज अधिक व्यवस्थित होती है।

9. Thesaurus निर्माण के चरणों का संक्षिप्त विवेचन कीजिए।

उत्तर: Thesaurus निर्माण में विषय क्षेत्र का चयन, शब्द संग्रह, शब्द विश्लेषण, संबंध निर्धारण, मानकीकरण तथा परीक्षण एवं संशोधन जैसे चरण शामिल होते हैं। इन चरणों से एक प्रभावी और उपयोगी controlled vocabulary tool विकसित होता है।

10. Folksonomy को अनुक्रमण भाषा के रूप में आलोचनात्मक रूप से विवेचित कीजिए।

उत्तर: Folksonomy उपयोगकर्ता-निर्मित टैगिंग पर आधारित होती है, जो लोकतांत्रिक और लचीली है।
हालाँकि, मानकीकरण के अभाव के कारण इसमें असंगति पाई जाती है, जिससे Precision प्रभावित होती है।
अतः इसे Controlled Vocabulary का पूरक माना जाता है, विकल्प नहीं।

2.18 लघु प्रश्न (Short Questions)

1. **Pre-coordination क्या है?**
   → विषय को पहले से निश्चित क्रम में एक ही शीर्षक के रूप में व्यक्त करना।
2. **BT और NT में अंतर बताइए।**
   → BT व्यापक शब्द है, NT उससे संकीर्ण शब्द होता है।
3. **UF का अर्थ स्पष्ट कीजिए।**
   → UF (Used For) समानार्थक या अप्रचलित शब्द को दर्शाता है।
4. **Subject Heading क्यों नियंत्रित शब्दावली है?**
   → क्योंकि इसमें केवल स्वीकृत और मानकीकृत शब्दों का प्रयोग होता है।
5. **Thesaurus में RT का महत्व क्या है?**
   → RT संबंधित अवधारणाओं को जोड़कर semantic clarity प्रदान करता है।
6. **Vocabulary Control से क्या अभिप्राय है?**
   → Vocabulary Control से तात्पर्य विषय-अभिव्यक्ति के लिए **मानकीकृत और स्वीकृत शब्दों** के चयन एवं प्रयोग से है, जिससे अनुक्रमण में **एकरूपता और स्पष्टता** बनी रहती है।
7. **Controlled Vocabulary Precision को कैसे बढ़ाती है?**
   → Controlled Vocabulary केवल **एक ही स्वीकृत शब्द** के प्रयोग की अनुमति देती है, जिससे **अप्रासंगिक परिणाम कम** होते हैं और खोज की **सटीकता (Precision)** बढ़ती है।

8. **Thesaurus में BT/NT संबंध क्यों आवश्यक हैं?**

→ BT/NT संबंध विषयों के बीच **क्रमिक (hierarchical) संबंध** स्थापित करते हैं, जिससे उपयोगकर्ता व्यापक से संकीर्ण विषय तक **सटीक खोज** कर सकता है।

9. Subject Heading और Keywords में एक अंतर लिखिए।

→ Subject Heading **नियंत्रित और मानकीकृत** शब्दावली पर आधारित होता है, जबकि Keywords **प्राकृतिक भाषा** के शब्द होते हैं जिनमें मानकीकरण नहीं होता।

10. Metadata और Indexing Language में क्या संबंध है?

→ Indexing Language, Metadata के **Subject या Descriptor तत्व** के रूप में कार्य करती है और डिजिटल संसाधनों की **विषय-पहचान** सुनिश्चित करती है।

11. Folksonomy क्यों असंगत मानी जाती है?

→ Folksonomy उपयोगकर्ता-निर्मित टैगिंग पर आधारित होती है, जहाँ एक ही विषय के लिए **अलग-अलग शब्द** प्रयोग किए जाते हैं, जिससे **असंगति** उत्पन्न होती है।

12. Thesaurus का एक प्रमुख उपयोग क्षेत्र लिखिए।

→ Thesaurus का प्रमुख उपयोग **डिजिटल पुस्तकालयों और ऑनलाइन सूचना पुनःप्राप्ति प्रणालियों** में semantic search को बेहतर बनाने के लिए किया जाता है।

2.19 बहुविकल्पीय प्रश्न (MCQs)

1. Thesaurus में UF का अर्थ है—
   A. Used File
   B. Used For
   C. User Function
   D. Uniform Form
   ✅उत्तर: **B**

2. Subject Heading की प्रकृति होती है—
   A. Post-coordinated
   B. Free-text
   C. Pre-coordinated
   D. Folksonomy-based
   ✅उत्तर: **C**

3. Thesaurus मुख्यतः किस पर आधारित होता है?
   A. Documents
   B. Authors
   C. Concepts

D. Formats

✅उत्तर: **C**

4. BT का अर्थ है—
A. Broader Term
B. Basic Term
C. Binary Term
D. Base Term

✅उत्तर: **A**

5. Thesaurus का प्रमुख लाभ है—
A. Shelf arrangement
B. Semantic relationships
C. Physical location
D. Accession control

✅उत्तर: **B**

6. Subject Heading Lists अधिक उपयुक्त हैं—
A. सोशल मीडिया के लिए
B. वेब टैगिंग के लिए
C. कैटलॉग और OPAC के लिए
D. ब्लॉग पोस्ट के लिए

✅उत्तर: **C**

7. "AI" को "Artificial Intelligence" से जोड़ना दर्शाता है—
A. RT
B. BT
C. NT
D. UF

✅उत्तर: **D**

8. Folksonomy आधारित होती है—
A. विशेषज्ञ नियंत्रण पर
B. उपयोगकर्ता-निर्मित टैगिंग पर
C. वर्गीकरण पर
D. कैटलॉग नियमों पर

✅उत्तर: **B**

9. Thesaurus में hierarchy किससे बनती है?
A. UF/RT
B. BT/NT
C. Scope notes

D. Tags

✅उत्तर: **B**

10. Recall बढ़ाने में अधिक सहायक है—
    A. Classification
    B. Subject Heading
    C. Thesaurus
    D. Shelf list

    ✅उत्तर: **C**

11. Controlled Vocabulary का मुख्य उद्देश्य है—
    A. अधिक शब्द जोड़ना
    B. विषय-अभिव्यक्ति में एकरूपता
    C. यूज़र टैगिंग
    D. फ्री-टेक्स्ट सर्च

    ✅उत्तर: **B**

12. Thesaurus किस प्रकार की अनुक्रमण भाषा है?
    A. Free-text
    B. Controlled
    C. Uncontrolled
    D. Folksonomy

    ✅उत्तर: **B**

13. Subject Heading Lists मुख्यतः उपयोग की जाती हैं—
    A. सोशल मीडिया में
    B. डिजिटल टैगिंग में
    C. Library Catalog में
    D. ब्लॉग लेखन में

    ✅उत्तर: **C**

14. "Global Warming" यदि UF है, तो PT होगा—
    A. Climate
    B. Environment
    C. Climate Change
    D. Sustainability

    ✅उत्तर: **C**

15. BT और NT संबंध दर्शाते हैं—
    A. समानार्थकता
    B. Hierarchy
    C. Homonym

D. Tagging

✅उत्तर: **B**

16. Folksonomy आधारित होती है—

A. विशेषज्ञ नियंत्रण पर

B. मानकीकृत शब्दावली पर

C. उपयोगकर्ता-निर्मित टैग पर

D. वर्गीकरण योजना पर

✅उत्तर: **C**

17. Subject Heading की सबसे बड़ी सीमा है—

A. उच्च Recall

B. Semantic linking

C. सीमित लचीलापन

D. User orientation

✅उत्तर: **C**

18. Thesaurus किसमें अधिक सहायक है?

A. Shelf arrangement

B. Semantic search

C. Accession register

D. Classification only

✅उत्तर: **B**

19. Metadata का मुख्य कार्य है—

A. दस्तावेज़ लिखना

B. दस्तावेज़ का वर्णन

C. टैग बनाना

D. अनुक्रमण हटाना

✅उत्तर: **B**

20. Recall बढ़ाने में सबसे अधिक सहायक है—

A. Subject Heading

B. Classification

C. Thesaurus

D. Shelf list

✅उत्तर: **C**

2.20 संदर्भ (References)

1. Lancaster, F. W. Indexing and Abstracting in Theory and Practice.

2. Foskett, A. C. The Subject Approach to Information.
3. Chowdhury, G. G. Introduction to Modern Information Retrieval.
4. Svenonius, E. The Intellectual Foundation of Information Organization.
5. Taylor, A. G. Introduction to Cataloging and Classification.
6. Aitchison, J., Gilchrist, A., & Bawden, D. Thesaurus Construction and Use.
7. ISO 25964-1/2. Thesauri and Interoperability with other Vocabularies.
8. Hjørland, B. Information Organization and Knowledge Organization.
9. Broughton, V. Essential Thesaurus Construction.
10. Soergel, D. Organizing Information.
11. National Information Standards Organization (NISO) guidelines (controlled vocabularies).
12. Library of Congress. Subject Headings (LCSH) documentation.
13. National Library of Medicine. MeSH documentation.
14. Manning, C. et al. Introduction to Information Retrieval.
15. रंगनाथन, एस. आर. – ज्ञान संगठन / वर्गीकरण सिद्धांत।
16. आर. एन. मिश्रा – ज्ञान संगठन एवं अनुक्रमण।
17. एस. डी. शर्मा – पुस्तकालय एवं सूचना विज्ञान।
18. पी. एस. जी. कुमार – सूचना पुनःप्राप्ति प्रणाली।
19. बी. के. सिन्हा – आधुनिक पुस्तकालय विज्ञान।
20. डॉ. के. के. वर्मा – पुस्तकालय संगठन एवं प्रबंधन।
21. डॉ. ए. के. सिंह – डिजिटल पुस्तकालय एवं मेटाडाटा।
22. डॉ. आर. पी. गुप्ता – अनुक्रमण, थिसॉरस एवं विषय अभिव्यक्ति।

इकाई–3 : अनुक्रमण भाषाएँ—वर्गीकरण योजनाएँ (Classification Schemes)

इकाई की रूपरेखा

3.1 प्रस्तावना
3.2 उद्देश्य
3.3 वर्गीकरण: अर्थ, उद्देश्य एवं क्षेत्र
3.4 पुस्तकालय वर्गीकरण की परिभाषाएँ
3.5 वर्गीकरण युग/विकास **(Evolution of Classification)**
3.6 वर्गीकरण का मूल्य एवं महत्व **(Value in Retrieval & Browsing)**
3.7 वर्गीकरण की विशेषताएँ एवं सिद्धांत **(Principles/Canons)**
3.8 वर्गीकरण में विषय विश्लेषण **(Subject Analysis)** की भूमिका
3.9 प्रमुख वर्गीकरण योजनाएँ—परिचय
3.9.1 DDC: संरचना, अंकन, उपयोग
3.9.2 UDC: विशेषताएँ एवं अनुप्रयोग
3.9.3 CC: Facet Analysis, PMEST, Notation
3.10 शेल्फ व्यवस्था एवं ब्राउज़िंग में वर्गीकरण
3.11 वर्गीकरण और अनुक्रमण का संबंध
3.12 डिजिटल/स्वचालित वर्गीकरण **(Automated Classification)** की अवधारणा
3.13 वर्गीकरण में चुनौतियाँ**: Multidisciplinary Subjects, Bias, Updates**
3.14 उदाहरण/केस: एक ही विषय को **DDC/UDC/CC** में कैसे रखा जाता है
3.15 सारांश
3.16 शब्दावली
3.17 निबंधात्मक प्रश्न एवं उत्तर
3.18 लघु प्रश्न **(Short Questions)**
3.19 बहुविकल्पीय प्रश्न **(MCQs)**
3.20 संदर्भ ग्रंथ सूची

3.1 प्रस्तावना (Introduction)

वर्गीकरण (Classification) पुस्तकालय एवं सूचना विज्ञान की वह **मौलिक बौद्धिक प्रक्रिया** है जिसके माध्यम से ज्ञान को **तार्किक, क्रमबद्ध और विषयानुसार समूहों** में संगठित किया जाता है। किसी भी पुस्तकालय या सूचना केंद्र में दस्तावेज़ों की संख्या निरंतर बढ़ती रहती है; ऐसे में यदि उन्हें किसी **मानकीकृत वर्गीकरण योजना** के अंतर्गत व्यवस्थित न किया जाए, तो न तो **सूचना पुनःप्राप्ति (Information Retrieval)** प्रभावी होगी और न ही **ब्राउज़िंग (Browsing)** संभव।

पारंपरिक पुस्तकालयों में वर्गीकरण का उद्देश्य शेल्फ पर पुस्तकों को सही क्रम में रखना था, जबकि **डिजिटल और स्वचालित परिवेश** में वर्गीकरण का लक्ष्य **डेटाबेस, OPAC और डिजिटल रिपॉज़िटरी** में खोज को सटीक बनाना है। इस प्रकार, वर्गीकरण योजनाएँ आज भी उतनी ही प्रासंगिक हैं जितनी कि पारंपरिक युग में थीं।

3.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात शिक्षार्थी—

- वर्गीकरण की अवधारणा, उद्देश्य एवं क्षेत्र को समझ सकेंगे
- पुस्तकालय वर्गीकरण की प्रमुख परिभाषाओं का विवेचन कर सकेंगे
- वर्गीकरण के विकास (Evolution) को ऐतिहासिक परिप्रेक्ष्य में समझेंगे
- DDC, UDC एवं CC जैसी प्रमुख वर्गीकरण योजनाओं की संरचना एवं उपयोग जानेंगे
- विषय विश्लेषण (Subject Analysis) की भूमिका स्पष्ट रूप से समझ सकेंगे
- एक ही विषय को विभिन्न वर्गीकरण योजनाओं में वर्गीकृत कर सकेंगे

3.2A अपेक्षित अधिगम परिणाम (Learning Outcomes)

इस इकाई के अध्ययन के पश्चात शिक्षार्थी—

- वर्गीकरण को एक **Indexing Language** के रूप में विश्लेषित कर सकेंगे
- DDC, UDC एवं CC के बीच **तुलनात्मक अंतर** स्पष्ट कर सकेंगे
- किसी दस्तावेज़ के लिए **उपयुक्त वर्ग संख्या** निर्धारित कर सकेंगे
- डिजिटल परिवेश में वर्गीकरण की सीमाओं एवं संभावनाओं का मूल्यांकन कर सकेंगे

3.3 वर्गीकरण: अर्थ, उद्देश्य एवं क्षेत्र

(A) वर्गीकरण का अर्थ (Meaning of Classification)

वर्गीकरण का शाब्दिक अर्थ है—

"समान गुणों या विशेषताओं के आधार पर वस्तुओं को समूहों में विभाजित करना।"
पुस्तकालय संदर्भ में वर्गीकरण का अर्थ है—
ज्ञान या दस्तावेज़ों को उनके विषयवस्तु (Subject Content) के आधार पर व्यवस्थित करना।

(B) वर्गीकरण के उद्देश्य (Objectives)

1. दस्तावेज़ों को तार्किक क्रम में व्यवस्थित करना
2. उपयोगकर्ता को सूचना खोजने में सहायता प्रदान करना
3. शेल्फ व्यवस्था एवं ब्राउज़िंग को सरल बनाना
4. संग्रह के प्रबंधन को सुगम बनाना
5. मानकीकरण (Standardization) सुनिश्चित करना

(C) वर्गीकरण का क्षेत्र (Scope)

- पुस्तकालय (Libraries)
- अभिलेखागार (Archives)
- डिजिटल लाइब्रेरी
- वेब एवं ऑनलाइन डेटाबेस

3.4 पुस्तकालय वर्गीकरण की परिभाषाएँ

- **एस. आर. रंगनाथन:**

  > "वर्गीकरण ज्ञान को उसके विषयों के अनुसार समूहित करने की एक व्यवस्थित प्रक्रिया है।"

- **हेनरी ब्लिस (Henry Bliss):**

  > "Classification is the systematic arrangement of knowledge."

इन परिभाषाओं से स्पष्ट है कि वर्गीकरण का केंद्र **ज्ञान और विषय** है, न कि केवल भौतिक वस्तुएँ।

3.5 वर्गीकरण का विकास (Evolution of Classification)

वर्गीकरण का विकास विभिन्न चरणों में हुआ—

(A) प्रारंभिक युग

- विषयानुसार अलमारियाँ
- लेखक या आकार आधारित व्यवस्था

(B) Enumerative Era

- **DDC (Dewey Decimal Classification)**
- सभी विषयों की सूची पहले से तैयार

(C) Analytico–Synthetic Era

- **Colon Classification (CC)**
- विषय का विश्लेषण एवं संश्लेषण (Analysis & Synthesis)

(D) आधुनिक एवं डिजिटल युग

- UDC का विस्तार
- स्वचालित एवं मशीन-आधारित वर्गीकरण

3.6 वर्गीकरण का मूल्य एवं महत्व (Value in Retrieval & Browsing)

वर्गीकरण का महत्व निम्न बिंदुओं से स्पष्ट होता है—

- समान विषय की सामग्री एक साथ उपलब्ध
- उपयोगकर्ता को ब्राउज़िंग द्वारा खोज की सुविधा
- OPAC एवं डिजिटल खोज में सहायता
- संग्रह का संतुलित विकास

निष्कर्ष:

"Good classification leads to effective information retrieval."

3.7 वर्गीकरण की विशेषताएँ एवं सिद्धांत (Principles / Canons)

मुख्य सिद्धांत (Canons by Ranganathan)

- **Canon of Consistency:** समान विषय के लिए समान वर्ग संख्या
- **Canon of Hospitality:** नए विषयों को समायोजित करने की क्षमता
- **Canon of Helpful Sequence:** तार्किक एवं सहायक क्रम

ये सिद्धांत वर्गीकरण को **लचीला (Flexible)** एवं **स्थायी (Sustainable)** बनाते हैं।

3.8 वर्गीकरण में विषय विश्लेषण (Subject Analysis) की भूमिका

विषय विश्लेषण वह प्रक्रिया है जिसमें दस्तावेज़ के—

- मुख्य विषय (Main Subject)
- उपविषय (Sub-topics)
- समय, स्थान, प्रक्रिया

की पहचान की जाती है।

Subject Analysis = Foundation of Classification

3.9 प्रमुख वर्गीकरण योजनाएँ—परिचय

3.9.1 DDC (Dewey Decimal Classification)

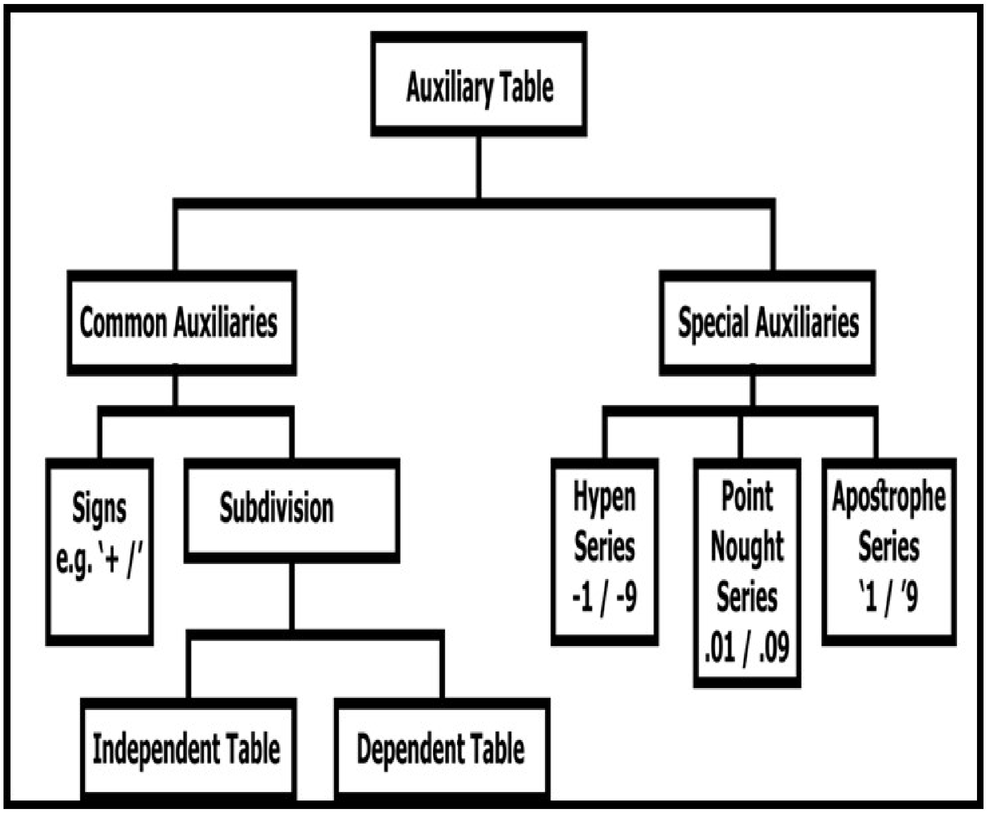
- **संरचना:** 10 मुख्य वर्ग (000–900)
- **Notation:** Pure Numeric
- **उपयोग:** सार्वजनिक एवं शैक्षणिक पुस्तकालय

001-099	Generalities	PREQUEL TO THE DEWEY DECIMAL SYSTEM	Encyclopedias, curiosities and wonders, unexplained mysteries
100-199	Philosophy	WHO AM I?	Books about the self, feelings, dreams, witchcraft
200-299	Religion	WHO MADE ME?	Christianity, Judaism, Buddhism, Hinduism, etc., and Mythology
300-399	Social Science	WHO'S THE GUY IN THE NEXT CAVE?	Customs, cultures, laws, manners, costumes, fairy tales
400-499	Languages	HOW DO I TALK TO THAT GUY?	Dictionaries, parts of speech, sign language, foreign language aids
500-599	Natural Science	LET'S TALK ABOUT THE WORLD WE SEE	Mathematics, earth, astronomy, chemistry, plants, wild animals
600-699	Applied Science	NOW LET'S MAKE STUFF OUT OF WHAT WE SEE	Inventions, robots, transportation, pets, recipe books
700-799	Arts and Recreation	NOW LET'S HAVE SOME FUN	Art, drawing, comics, handicrafts, music, games, sports
800-899	Literature	LET'S TELL OUR CHILDREN HOW WONDERFUL WE ARE	Poetry, plays, classic literature, jokes, riddles
900-999	Geography and History	LET'S TELL OUR FUTURE CHILDREN HOW WONDERFUL WE WERE	Landforms, travel, atlases, exploration, countries
92 and 920	Biography and Collective Biography	FIND OUT ABOUT FAMOUS PEOPLE	Single person: filed by last name of subject Multiple people: by author

उदाहरण:
Computer Science → **004**

3.9.2 UDC (Universal Decimal Classification)

- **विशेषताएँ:** Faceted structure
- **Notation:** Numeric + symbols
- **उपयोग:** अनुसंधान एवं विशेष पुस्तकालय



3.9.3 CC (Colon Classification): PMEST

रंगनाथन द्वारा प्रतिपादित **पाँच मौलिक श्रेणियाँ**—

- **P – Personality**
- **M – Matter**
- **E – Energy**
- **S – Space**
- **T – Time**

सरल उदाहरण:
"Digital Library Services in India (2020)"

- P: Digital Library
- E: Services
- S: India
- T: 2020

3.10 शेल्फ व्यवस्था एवं ब्राउज़िंग में वर्गीकरण

वर्गीकरण से समान विषय की पुस्तकें पास-पास रखी जाती हैं, जिससे उपयोगकर्ता **ब्राउज़िंग द्वारा खोज** कर सकता है—जो विशेषकर **ODL learners** के लिए अत्यंत उपयोगी है।

3.11 वर्गीकरण और अनुक्रमण का संबंध

वर्गीकरण	अनुक्रमण
व्यापक विषय व्यवस्था	विशिष्ट पहुँच बिंदु
शेल्फ आधारित	खोज आधारित
स्थायी	अधिक लचीला

दोनों मिलकर **Effective Retrieval System** बनाते हैं।

3.12 डिजिटल/स्वचालित वर्गीकरण (Automated Classification)

Automated classification में **Machine Learning classifiers**, **Text mining**, और **Ontology-based systems** का प्रयोग किया जाता है

- यह बड़े Digital Repositories, Institutional Repositories एवं Web-scale systems के लिए उपयोगी है
- हालांकि, **semantic ambiguity** और **training data bias** इसकी प्रमुख सीमाएँ हैं

3.13 वर्गीकरण में चुनौतियाँ

- Multidisciplinary Subjects
- सांस्कृतिक Bias
- निरंतर अद्यतन (Updates)

3.14 केस/उदाहरण

विषय: *Digital Libraries in India*

- **DDC:** 025.04
- **UDC:** 027.7:004(540)
- **CC:** P (Library) + E (Digital) + S (India)

3.15 सारांश (Summary)

इस इकाई में **वर्गीकरण योजनाओं** को अनुक्रमण भाषाओं के एक महत्वपूर्ण रूप के रूप में विस्तार से समझाया गया है। वर्गीकरण पुस्तकालय एवं सूचना विज्ञान की वह आधारभूत प्रक्रिया है, जिसके माध्यम से ज्ञान को **तार्किक, क्रमबद्ध और विषयानुसार** व्यवस्थित किया जाता है।

इकाई के प्रारंभ में वर्गीकरण के **अर्थ, उद्देश्य और क्षेत्र** पर प्रकाश डाला गया, जिससे यह स्पष्ट होता है कि वर्गीकरण केवल पुस्तकों को शेल्फ पर रखने की तकनीक नहीं है, बल्कि यह **सूचना पुनःप्राप्ति (Information Retrieval)** का एक प्रभावी साधन है।

इसके पश्चात पुस्तकालय वर्गीकरण की प्रमुख परिभाषाओं तथा **विकास क्रम (Evolution of Classification)** की चर्चा की गई, जिसमें Enumerative और Analytico–Synthetic वर्गीकरण का अंतर स्पष्ट किया गया।

इकाई का मुख्य भाग **DDC, UDC और Colon Classification (CC)** जैसी प्रमुख वर्गीकरण योजनाओं को समर्पित है। DDC की सरल संख्यात्मक संरचना, UDC की फेसटेड और सहायक तालिकाओं पर आधारित प्रणाली तथा CC की **PMEST अवधारणा** को उदाहरण सहित समझाया गया है।

इसके अतिरिक्त, वर्गीकरण में **विषय विश्लेषण (Subject Analysis)** की भूमिका, शेल्फ व्यवस्था एवं ब्राउज़िंग में वर्गीकरण का महत्व तथा **डिजिटल/स्वचालित वर्गीकरण** की आधुनिक अवधारणा को भी शामिल किया गया है।

अंततः, यह इकाई यह निष्कर्ष प्रस्तुत करती है कि **वर्गीकरण और अनुक्रमण एक-दूसरे के पूरक** हैं और दोनों के संयुक्त उपयोग से ही प्रभावी सूचना संगठन एवं पुनःप्राप्ति संभव है।

3.16 संदर्भ (References – In-text / Suggested Readings)

(यह भाग शिक्षार्थियों को आगे अध्ययन हेतु दिशा देता है)

- Dewey Decimal Classification का उपयोग सार्वजनिक पुस्तकालयों में
- UDC का प्रयोग विशेष एवं तकनीकी पुस्तकालयों में
- Colon Classification का अकादमिक एवं शोध स्तर पर महत्व
- डिजिटल लाइब्रेरी में वर्गीकरण और मेटाडाटा का संबंध

3.17 शब्दावली (Glossary)

1. **वर्गीकरण (Classification):** विषय के आधार पर ज्ञान को समूहों में व्यवस्थित करने की प्रक्रिया।
2. **अनुक्रमण भाषा (Indexing Language):** वह भाषा जो सूचना संसाधनों के विषय को अभिव्यक्त करती है।
3. **Notation (अंकन):** वर्गीकरण में प्रयुक्त संख्यात्मक/प्रतीकात्मक संकेत।
4. **Facet (फेसट)**: विषय का वह पहलू जो किसी विशिष्ट गुण को दर्शाता है।
5. **PMEST:** Colon Classification की पाँच मौलिक श्रेणियाँ—Personality, Matter, Energy, Space, Time।
6. **Browsing:** शेल्फ या डिजिटल संग्रह में विषयानुसार अवलोकन द्वारा खोज।
7. **Enumerative Classification:** पूर्व-सूचित विषयों की सूची पर आधारित वर्गीकरण।
8. **Analytico–Synthetic Classification:** विषय विश्लेषण एवं संश्लेषण पर आधारित वर्गीकरण।
9. **Knowledge Organization (ज्ञान संगठन)**: ज्ञान को व्यवस्थित करने की समग्र प्रक्रिया जिसमें वर्गीकरण, अनुक्रमण और वर्णन शामिल हैं।
10. **Relative Location**: वर्ग संख्या के आधार पर पुस्तकों का सापेक्ष स्थान निर्धारण।
11. **Auxiliary Tables (सहायक तालिकाएँ)**: UDC में प्रयुक्त तालिकाएँ जो विषय को विस्तारित करती हैं।
12. **Notation Economy**: कम प्रतीकों में अधिक विषय अभिव्यक्ति की क्षमता।
13. **Semantic Ambiguity**: एक ही शब्द के अनेक अर्थ होने की स्थिति।

3.18 निबंधात्मक प्रश्न (Essay Questions)

1. पुस्तकालय वर्गीकरण से आप क्या समझते हैं? इसके उद्देश्य एवं क्षेत्र स्पष्ट कीजिए।
2. वर्गीकरण के विकास (Evolution of Classification) का ऐतिहासिक विवेचन कीजिए।
3. DDC की संरचना, विशेषताएँ एवं सीमाओं का वर्णन कीजिए।
4. UDC को एक फेसटेड वर्गीकरण योजना के रूप में समझाइए।
5. Colon Classification की PMEST अवधारणा को उदाहरण सहित स्पष्ट कीजिए।
6. वर्गीकरण में विषय विश्लेषण की भूमिका पर चर्चा कीजिए।

7. शेल्फ व्यवस्था एवं ब्राउज़िंग में वर्गीकरण का महत्व स्पष्ट कीजिए।
8. वर्गीकरण और अनुक्रमण के बीच संबंध स्पष्ट कीजिए।
9. डिजिटल एवं स्वचालित वर्गीकरण की अवधारणा और उपयोग पर टिप्पणी कीजिए।
10. बहुविषयी (Multidisciplinary) विषयों के वर्गीकरण में आने वाली चुनौतियों की चर्चा कीजिए।

लघु प्रश्न (Short Questions – 10)

1. वर्गीकरण क्या है?
2. DDC में कितने मुख्य वर्ग होते हैं?
3. UDC की कोई दो विशेषताएँ लिखिए।
4. PMEST का पूरा नाम लिखिए।
5. Notation से क्या अभिप्राय है?
6. Browsing क्या है?
7. Enumerative Classification का अर्थ बताइए।
8. Subject Analysis क्यों आवश्यक है?
9. Canon of Consistency क्या है?
10. डिजिटल वर्गीकरण से क्या तात्पर्य है?

बहुविकल्पीय प्रश्न (MCQs – 10)

1. PMEST अवधारणा किस वर्गीकरण से संबंधित है?
   A. DDC
   B. UDC
   C. CC
   D. LCC
   ✔ उत्तर: C
2. DDC का विकास किसने किया?
   A. Ranganathan
   B. Dewey
   C. Bliss
   D. Cutter
   ✔ उत्तर: B
3. UDC मुख्यतः किस प्रकार की वर्गीकरण योजना है?
   A. Enumerative
   B. Alphabetical
   C. Faceted

D. Chronological
✔ उत्तर: C

4. Browsing का संबंध है—
A. अनुक्रमण से
B. वर्गीकरण से
C. संदर्भ सेवा से
D. अधिग्रहण से
✔ उत्तर: B

5. Canon of Hospitality किससे संबंधित है?
A. विषय विश्लेषण
B. विस्तारशीलता
C. शेल्फ व्यवस्था
D. अंकन
✔ उत्तर: B

3.19 अभ्यास प्रश्नों के उत्तर

A. निबंधात्मक प्रश्नों के उत्तर – संकेत (Essay Answers – Key Points)

प्रश्न 1. पुस्तकालय वर्गीकरण से आप क्या समझते हैं? इसके उद्देश्य एवं क्षेत्र स्पष्ट कीजिए।

उत्तर: पुस्तकालय वर्गीकरण (Library Classification) ज्ञान संगठन की वह बौद्धिक प्रक्रिया है, जिसके माध्यम से सूचना संसाधनों को उनके विषयवस्तु (Subject Content) के आधार पर तार्किक, क्रमबद्ध एवं व्यवस्थित समूहों में विभाजित किया जाता है। इसका मुख्य उद्देश्य यह सुनिश्चित करना है कि समान विषय से संबंधित सामग्री एक साथ उपलब्ध हो, जिससे सूचना की खोज, पुनःप्राप्ति एवं ब्राउज़िंग सरल हो सके।

उद्देश्य:
पुस्तकालय वर्गीकरण के प्रमुख उद्देश्य निम्नलिखित हैं—

- दस्तावेज़ों को विषयानुसार व्यवस्थित करना
- उपयोगकर्ताओं को सूचना खोजने में सहायता प्रदान करना
- शेल्फ व्यवस्था एवं ब्राउज़िंग को सरल बनाना
- संग्रह प्रबंधन को सुव्यवस्थित करना
- सूचना पुनःप्राप्ति प्रणाली को प्रभावी बनाना

क्षेत्र(Scope):
वर्गीकरण का प्रयोग केवल पारंपरिक पुस्तकालयों तक सीमित नहीं है, बल्कि यह—

- डिजिटल पुस्तकालय
- ऑनलाइन डेटाबेस
- अभिलेखागार
- सूचना केंद्र एवं रिपॉज़िटरी
  में भी समान रूप से उपयोगी है।

इस प्रकार, पुस्तकालय वर्गीकरण सूचना संगठन की आधारशिला है।

प्रश्न 2. वर्गीकरण के विकास (Evolution of Classification) का ऐतिहासिक विवेचन कीजिए।

उत्तर:वर्गीकरण का विकास ज्ञान के विस्तार के साथ क्रमिक रूप से हुआ है। इसे मुख्यतः चार चरणों में समझा जा सकता है—

(1)प्रारंभिकयुग:
इस युग में पुस्तकों को लेखक, आकार या विषय के सामान्य आधार पर रखा जाता था। कोई मानकीकृत प्रणाली उपलब्ध नहीं थी।

(2)EnumerativeEra:
इस युग में सभी संभावित विषयों की पूर्व-निर्धारित सूची बनाई गई। Dewey Decimal Classification (DDC) इसका प्रमुख उदाहरण है, जिसमें विषयों को पहले से सूचीबद्ध किया गया।

(3) Analytico–Synthetic Era:
इस चरण में विषयों का विश्लेषण (Analysis) एवं संश्लेषण (Synthesis) किया गया। Colon Classification (CC) इस युग की प्रतिनिधि योजना है, जिसमें PMEST अवधारणा प्रस्तुत की गई।

(4) आधुनिक एवं डिजिटल युग:
इस युग में UDC का विस्तार हुआ तथा स्वचालित एवं मशीन आधारित वर्गीकरण का विकास हुआ।

इस प्रकार, वर्गीकरण का विकास स्थिर न होकर सतत और प्रगतिशील रहा है।

प्रश्न 3. DDC की संरचना, विशेषताएँ एवं सीमाओं का वर्णन कीजिए।

उत्तर: **Dewey Decimal Classification (DDC)** का विकास Melvil Dewey द्वारा किया गया। यह विश्व की सर्वाधिक प्रचलित वर्गीकरण योजना है।

संरचना:
DDC में ज्ञान को 10 मुख्य वर्गों (000–900) में विभाजित किया गया है। प्रत्येक मुख्य वर्ग को आगे उपवर्गों एवं उपविभागों में विभाजित किया गया है।

विशेषताएँ:

- Pure Numeric Notation
- सरल एवं तार्किक संरचना
- व्यापक अंतरराष्ट्रीय उपयोग
- सार्वजनिक एवं शैक्षणिक पुस्तकालयों के लिए उपयुक्त

सीमाएँ:

- पश्चिमी ज्ञान पर अधिक केंद्रित
- नवीन एवं बहुविषयी विषयों के लिए सीमित लचीलापन
- अत्यधिक विस्तार में कठिनाई

निष्कर्षतः, DDC सरल एवं व्यावहारिक होते हुए भी कुछ सीमाओं से युक्त है।

प्रश्न 4. UDC को एक Faceted वर्गीकरण योजना के रूप में समझाइए।

उत्तर: Universal Decimal Classification (UDC) एक उन्नत एवं लचीली वर्गीकरण योजना है, जो मुख्यतः Faceted सिद्धांत पर आधारित है।

मुख्य विशेषताएँ:

- विषयों को विभिन्न पहलुओं (Facets) में विभाजित करने की सुविधा
- सहायक तालिकाओं (Auxiliary Tables) का प्रयोग
- संख्यात्मक एवं प्रतीकात्मक अंकन
- जटिल एवं बहुविषयी विषयों के लिए उपयुक्त

UDC विशेष पुस्तकालयों, अनुसंधान संस्थानों एवं तकनीकी संग्रहों में अधिक उपयोगी है। इसकी Faceted संरचना इसे आधुनिक सूचना परिवेश के लिए उपयुक्त बनाती है।

प्रश्न 5. Colon Classification की PMEST अवधारणा को उदाहरण सहित स्पष्ट कीजिए।

उत्तर: Colon Classification (CC) का प्रतिपादन डॉ. एस. आर. रंगनाथन द्वारा किया गया। यह Analytico–Synthetic वर्गीकरण योजना है।

PMEST पाँच मौलिक श्रेणियों का प्रतिनिधित्व करता है—

- **P – Personality**
- **M – Matter**
- **E – Energy**
- **S – Space**
- **T – Time**

उदाहरण:
"Digital Library Services in India (2020)"

- P: Digital Library
- E: Services
- S: India
- T: 2020

PMEST अवधारणा जटिल विषयों के सटीक वर्गीकरण में अत्यंत सहायक है।

प्रश्न 6. वर्गीकरण में विषय विश्लेषण (Subject Analysis) की भूमिका पर चर्चा कीजिए।

उत्तर: विषय विश्लेषण वर्गीकरण की आधारभूत प्रक्रिया है। इसके बिना सही वर्ग संख्या निर्धारित नहीं की जा सकती।

इस प्रक्रिया में—

- मुख्य विषय की पहचान
- उपविषयों का निर्धारण
- स्थान, समय एवं प्रक्रिया का विश्लेषण किया जाता है।

सटीक विषय विश्लेषण से ही प्रभावी वर्गीकरण संभव है। अतः कहा जा सकता है कि— **"Subject Analysis is the foundation of Classification."**

प्रश्न 7. शेल्फ व्यवस्था एवं ब्राउज़िंग में वर्गीकरण का महत्व स्पष्ट कीजिए।

उत्तर: वर्गीकरण शेल्फ व्यवस्था का मूल आधार है। इसके माध्यम से समान विषय की पुस्तकें एक साथ रखी जाती हैं।

महत्व:

- उपयोगकर्ता को ब्राउज़िंग द्वारा खोज की सुविधा
- स्वयं अध्ययन (Self-learning) को प्रोत्साहन
- संग्रह का तार्किक एवं सुव्यवस्थित स्वरूप

ODL प्रणाली में ब्राउज़िंग आधारित खोज अत्यंत उपयोगी सिद्ध होती है।

प्रश्न 8. वर्गीकरण और अनुक्रमण के बीच संबंध स्पष्ट कीजिए।

उत्तर: वर्गीकरण और अनुक्रमण दोनों सूचना संगठन की पूरक प्रक्रियाएँ हैं।

- वर्गीकरण विषयों की **व्यापक व्यवस्था** प्रदान करता है
- अनुक्रमण **विशिष्ट पहुँच बिंदु** उपलब्ध कराता है

वर्गीकरण शेल्फ आधारित होता है, जबकि अनुक्रमण खोज आधारित। दोनों के संयुक्त उपयोग से प्रभावी सूचना पुनःप्राप्ति संभव होती है।

प्रश्न 9. डिजिटल एवं स्वचालित वर्गीकरण की अवधारणा एवं उपयोग पर टिप्पणी कीजिए।

उत्तर: डिजिटल युग में स्वचालित वर्गीकरण का महत्व तेजी से बढ़ा है। इसमें मशीन लर्निंग, मेटाडाटा एवं एल्गोरिद्म का प्रयोग किया जाता है।

उपयोग:

- डिजिटल पुस्तकालय
- संस्थागत रिपॉज़िटरी
- बड़े ऑनलाइन डेटाबेस

हालाँकि यह समय बचाता है, परंतु semantic ambiguity एवं bias इसकी सीमाएँ हैं।

प्रश्न 10. बहुविषयी (Multidisciplinary) विषयों के वर्गीकरण में आने वाली चुनौतियों की चर्चा कीजिए।

उत्तर: बहुविषयी विषय एक से अधिक विषय क्षेत्रों से संबंधित होते हैं, जिससे वर्गीकरण जटिल हो जाता है।

मुख्य चुनौतियाँ:

- उचित मुख्य विषय का चयन
- विभिन्न Facets का संतुलन
- वर्गीकरण योजनाओं की सीमाएँ
- निरंतर अद्यतन की आवश्यकता

इन चुनौतियों के समाधान हेतु Faceted एवं लचीली योजनाएँ अधिक उपयुक्त मानी जाती हैं।

B. लघु प्रश्नों के उत्तर (Short Answers)

1. **वर्गीकरण:** विषय के आधार पर ज्ञान का व्यवस्थित संगठन।
2. **DDC:** Dewey Decimal Classification; 10 मुख्य वर्ग।
3. **UDC की दो विशेषताएँ:** Faceted structure, auxiliary tables।
4. **PMEST:** Personality, Matter, Energy, Space, Time।
5. **Notation:** वर्गीकरण में प्रयुक्त संख्यात्मक/प्रतीकात्मक संकेत।
6. **Browsing:** विषयानुसार शेल्फ/डिजिटल अवलोकन द्वारा खोज।
7. **Enumerative Classification:** पूर्व-सूचित विषयों की सूची पर आधारित।
8. **Subject Analysis:** दस्तावेज़ के विषय घटकों की पहचान।
9. **Canon of Consistency:** समान विषय के लिए समान वर्ग संख्या।
10. **डिजिटल वर्गीकरण:** कंप्यूटर/एल्गोरिद्म आधारित वर्गीकरण।

C. MCQs के उत्तर (Answer Key)

1. PMEST अवधारणा संबंधित है — **CC**
2. DDC के जनक — **Melvil Dewey**
3. UDC किस प्रकार की योजना है — **Faceted**
4. Browsing का संबंध — **वर्गीकरण**
5. Canon of Hospitality — **नए विषयों को समायोजित करने की क्षमता**
6. Pure numeric notation किसमें है — **DDC**
7. Analytico–Synthetic योजना — **Colon Classification**
8. Subject analysis का उद्देश्य — **सही वर्ग संख्या बनाना**
9. Facet का अर्थ — **विषय का पहलू**
10. Automated classification आधारित है — **Metadata एवं Algorithms**

3.20 संदर्भ ग्रंथ सूची (Bibliography)

1. Ranganathan, S. R. Colon Classification.
2. Dewey, M. Dewey Decimal Classification and Relative Index.

3. Chowdhury, G. G. Knowledge Organization.
4. Bliss, H. E. The Organization of Knowledge.
5. Mills, J. A Modern Outline of Classification.
6. Satija, M. P. Colon Classification.
7. Hunter, E. J. Classification Made Simple.
8. रंगनाथन, एस. आर. – कोलन वर्गीकरण।
9. पी. एस. जी. कुमार – ज्ञान संगठन एवं सूचना पुनःप्राप्ति।
10. आर. एन. मिश्रा – पुस्तकालय वर्गीकरण।
11. के. के. वर्मा – पुस्तकालय एवं सूचना विज्ञान के सिद्धांत।
12. डॉ. एस. डी. शर्मा – आधुनिक पुस्तकालय प्रणाली।
13. डॉ. ए. के. सिंह – ज्ञान संगठन एवं डिजिटल पुस्तकालय।

इकाई–4 : अनुक्रमण प्रणाली और तकनीक (Indexing Systems & Techniques)

इकाई की रूपरेखा

4.1 प्रस्तावना
4.2 उद्देश्य
4.3 अनुक्रमण प्रणाली: अर्थ एवं विकास
4.4 अनुक्रमण के उद्देश्य एवं कार्य
4.5 अनुक्रमण प्रणालियों के प्रकार
4.6 अनुक्रमण तकनीकें
4.7 Abstracting & Indexing सेवाएँ
4.8 स्वचालित/मशीनी अनुक्रमण
4.9 मेटाडाटा आधारित अनुक्रमण
4.10 वेब अनुक्रमण
4.11 गुणवत्ता नियंत्रण
4.12 व्यावहारिक उदाहरण
4.13 आधुनिक प्रवृत्तियाँ
4.14 सामान्य त्रुटियाँ
4.15 सारांश
4.16 शब्दावली
4.17 निबंधात्मक प्रश्न एवं उत्तर
4.18 लघु प्रश्न (Short Questions)
4.19 बहुविकल्पीय प्रश्न (MCQs)
4.20 संदर्भ ग्रंथ सूची

4.1 प्रस्तावना (Introduction)

अनुक्रमण प्रणाली पुस्तकालय एवं सूचना विज्ञान के क्षेत्र में सूचना पुनःप्राप्ति (Information Retrieval) की **मौलिक और अनिवार्य प्रक्रिया** है। किसी भी पुस्तकालय, सूचना केन्द्र, डिजिटल रिपॉजिटरी या वेब-आधारित सूचना प्रणाली का वास्तविक मूल्य तभी सिद्ध होता है जब उपयोगकर्ता अपनी सूचना आवश्यकता के अनुसार **सटीक, प्रासंगिक और समयानुकूल सूचना** प्राप्त कर सके।

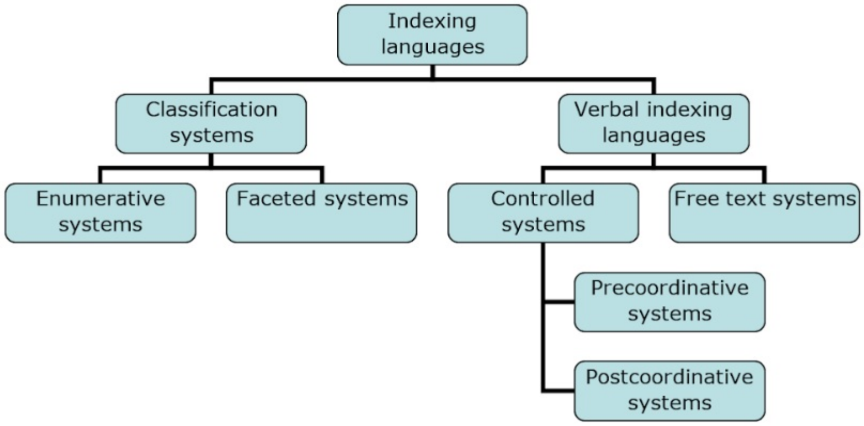
सूचना के तीव्र विस्तार के वर्तमान युग में केवल संसाधनों का संग्रह पर्याप्त नहीं है। आवश्यकता इस बात की है कि इन संसाधनों को इस प्रकार व्यवस्थित किया जाए कि वे **खोजयोग्य (discoverable)** और **उपयोगी (usable)** बन सकें। यही कार्य अनुक्रमण प्रणाली करती है।

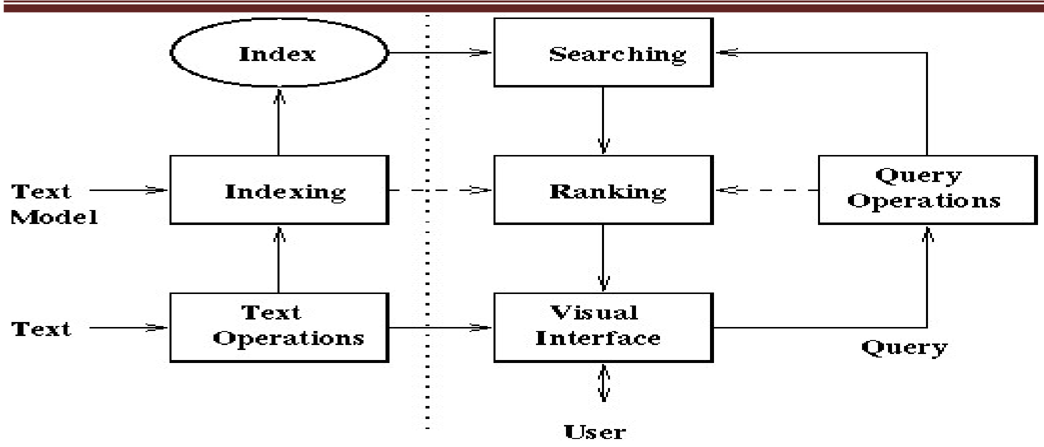
अनुक्रमण प्रणाली दस्तावेज़ की विषय-वस्तु का **बौद्धिक विश्लेषण** कर उसे ऐसे अनुक्रमण पदों (Indexing Terms) में रूपांतरित करती है, जो उपयोगकर्ता की खोज-भाषा और सूचना आवश्यकता से मेल खाते हों। इस प्रकार अनुक्रमण प्रणाली दस्तावेज़ और उपयोगकर्ता के बीच **बौद्धिक सेतु (Intellectual Bridge)** का कार्य करती है।

ODL संदर्भ में महत्व

Open & Distance Learning प्रणाली में शिक्षार्थी स्वयं अध्ययन करता है। ऐसे में अनुक्रमण प्रणाली का महत्व और बढ़ जाता है क्योंकि:

- शिक्षार्थी स्वयं सूचना खोजता है
- डिजिटल संसाधनों पर निर्भरता अधिक होती है
- सटीक अनुक्रमण बिना शिक्षक की सहायता के सीखने को संभव बनाता है





4.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात शिक्षार्थी निम्नलिखित योग्यताएँ प्राप्त कर सकेगा/सकेगी:

1. अनुक्रमण प्रणाली की अवधारणा और आवश्यकता को स्पष्ट रूप से समझ सकेगा
2. अनुक्रमण प्रणालियों के विकास और परिवर्तन का विश्लेषण कर सकेगा
3. विभिन्न अनुक्रमण प्रणालियों एवं तकनीकों में अंतर कर सकेगा
4. डिजिटल, वेब और स्वचालित अनुक्रमण की भूमिका को समझ सकेगा
5. किसी दस्तावेज़ या लेख का व्यावहारिक अनुक्रमण कर सकेगा

4.3 अनुक्रमण प्रणाली: अर्थ एवं विकास

अर्थ (Meaning of Indexing System)

अनुक्रमण प्रणाली से तात्पर्य उस संगठित व्यवस्था से है जिसके अंतर्गत दस्तावेज़ों की विषय-वस्तु का विश्लेषण कर उसे नियंत्रित और मानकीकृत पदों में व्यक्त किया जाता है, ताकि सूचना की खोज सरल, त्वरित और सटीक हो सके।

सरल शब्दों में,
अनुक्रमण प्रणाली वह प्रक्रिया है जो "दस्तावेज़" को "खोजयोग्य सूचना" में बदलती है।

विकास (Evolution of Indexing Systems)

अनुक्रमण प्रणाली का विकास सूचना की मात्रा, स्वरूप और उपयोग की आवश्यकता के अनुसार हुआ है।

1. प्रारंभिक चरण

- पुस्तकों के अंत में विषय-सूचियाँ
- मैनुअल अनुक्रम (Printed Indexes)
- सीमित संग्रह और सीमित उपयोगकर्ता

2. मानकीकरण चरण

- नियंत्रित शब्दावली का विकास
- विषय शीर्षकसूचियाँ
- पुस्तकालयों में एकरूपता की स्थापना

3. कंप्यूटरीकरण चरण

- OPAC (Online Public Access Catalogue)
- बिब्लियोग्राफिक डेटाबेस
- मशीन-पठनीय रिकॉर्ड

4. डिजिटल एवं वेब चरण

- डिजिटल लाइब्रेरी
- वेब आधारित अनुक्रमण
- सर्च इंजन इंडेक्स

5. आधुनिक (Semantic) चरण

- Artificial Intelligence
- Natural Language Processing
- Semantic & Concept-based Indexing

यह विकास दर्शाता है कि अनुक्रमण प्रणाली अब केवल संगठन का साधन नहीं रही, बल्कि **ज्ञान विश्लेषण और ज्ञान खोज** का प्रमुख उपकरण बन चुकी है।

4.4 अनुक्रमण के उद्देश्य एवं कार्य

4.4.1 अनुक्रमण के उद्देश्य

अनुक्रमण के प्रमुख उद्देश्य निम्नलिखित हैं:

1. उपयोगकर्ता को सटीक और प्रासंगिक सूचना प्रदान करना
2. विशाल सूचना संग्रह पर बौद्धिक नियंत्रण स्थापित करना
3. शोध, शिक्षण और अध्ययन को सुगम बनाना
4. सूचना संसाधनों के उपयोग को अधिकतम करना

4.4.2 अनुक्रमण के प्रमुख कार्य

अनुक्रमण प्रणाली के अंतर्गत निम्नलिखित कार्य किए जाते हैं:

- विषय विश्लेषण (Subject Analysis)
- अवधारणाओं की पहचान
- उपयुक्त अनुक्रमण पदों का चयन
- पदों का मानकीकरण
- क्रॉस-रेफरेंस का निर्माण

शैक्षणिक उदाहरण

यदि किसी लेख का विषय है
"Digital Libraries and Open Access in Indian Universities"
तो अनुक्रमण प्रणाली द्वारा मुख्य अवधारणाएँ होंगी:

- Digital Libraries
- Open Access
- Indian Universities

4.5 अनुक्रमण प्रणालियों के प्रकार (Types of Indexing Systems)

अनुक्रमण प्रणालियों को मुख्यतः दो व्यापक वर्गों में विभाजित किया जाता है—

1. **प्री-कोऑर्डिनेट अनुक्रमण प्रणाली (Pre-coordinate Indexing)**
2. **पोस्ट-कोऑर्डिनेट अनुक्रमण प्रणाली (Post-coordinate Indexing)**

यह वर्गीकरण इस आधार पर किया जाता है कि **अनुक्रमण पदों का संयोजन किस चरण में किया जाता है**—अनुक्रमण के समय या खोज के समय।

4.5.1 प्री-कोऑर्डिनेट अनुक्रमण प्रणाली (Pre-coordinate Indexing)

अवधारणा

प्री-कोऑर्डिनेट अनुक्रमण प्रणाली वह प्रणाली है जिसमें किसी दस्तावेज़ की विषय-वस्तु से संबंधित अनुक्रमण पदों को **अनुक्रमण के समय ही संयोजित (co-ordinate)** कर दिया जाता है। अर्थात्, दस्तावेज़ का विश्लेषण करने वाला अनुक्रमणकर्ता पहले से ही यह निर्णय ले लेता है कि विभिन्न अवधारणाओं को किस क्रम और रूप में प्रस्तुत किया जाएगा।

परिभाषात्मक रूप में

प्री-कोऑर्डिनेट अनुक्रमण वह प्रणाली है जिसमें विषय की विभिन्न अवधारणाओं को एक संयुक्त अनुक्रमण पद के रूप में पूर्व-निर्धारित किया जाता है।

उदाहरण

यदि किसी दस्तावेज़ का विषय है—
"Circulation of Periodicals in University Libraries"

तो प्री-कोऑर्डिनेट अनुक्रमण में इसे इस प्रकार प्रस्तुत किया जा सकता है:
Periodicals — Circulation — University Libraries

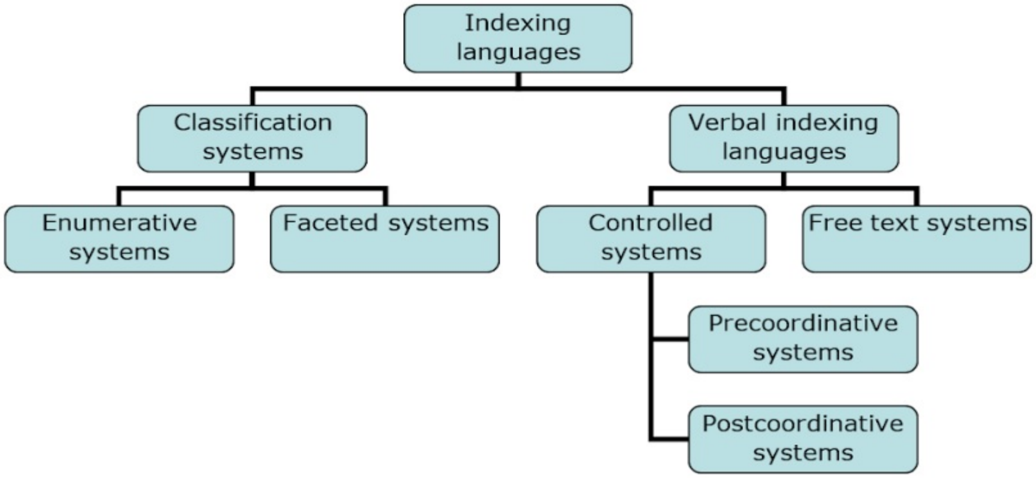
यहाँ तीनों अवधारणाएँ (Periodicals, Circulation, University Libraries) पहले ही एक निश्चित क्रम में संयोजित कर दी गई हैं।

विशेषताएँ

- अनुक्रमण पदों का संयोजन पहले से किया जाता है
- नियंत्रित शब्दावली का प्रयोग
- खोज में अधिक सटीकता (High Precision)
- पारंपरिक पुस्तकालय प्रणालियों के लिए उपयुक्त

सीमाएँ

- खोज में लचीलापन कम
- जटिल विषयों के लिए लंबी विषय-संरचनाएँ
- उपयोगकर्ता की बदलती खोज-भाषा के अनुकूलन में कठिनाई



4.5.2 पोस्ट-कोऑर्डिनेट अनुक्रमण प्रणाली (Post-coordinate Indexing)

अवधारणा

पोस्ट-कोऑर्डिनेट अनुक्रमण प्रणाली वह प्रणाली है जिसमें अनुक्रमण पदों को **स्वतंत्र रूप से** रखा जाता है और उनका संयोजन **खोज के समय उपयोगकर्ता द्वारा** किया जाता है। इसमें अनुक्रमणकर्ता पदों को अलग-अलग दर्ज करता है, जबकि खोज के समय उपयोगकर्ता अपनी आवश्यकता के अनुसार उन्हें जोड़ता है।

उदाहरण

यदि विषय है—
"Digital Libraries in India"

तो पोस्ट-कोऑर्डिनेट अनुक्रमण में अनुक्रमण पद होंगे:

- Digital Libraries
- India

खोज के समय उपयोगकर्ता इन्हें जोड़कर परिणाम प्राप्त करता है।

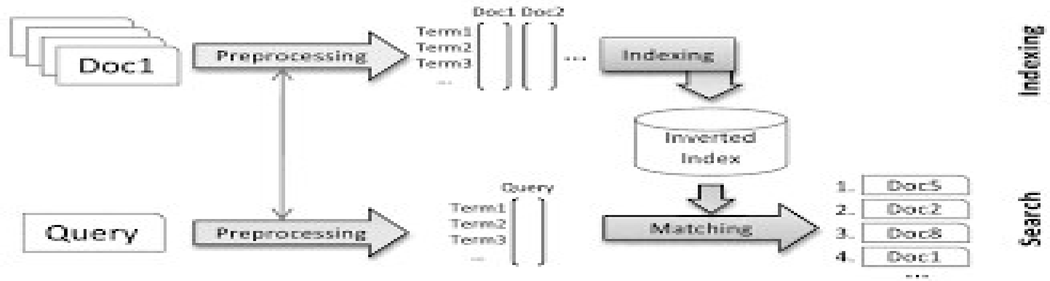
विशेषताएँ

- खोज में अधिक लचीलापन
- Recall अधिक (अधिक परिणाम)

- डिजिटल डेटाबेस और ऑनलाइन खोज प्रणालियों के लिए उपयुक्त

सीमाएँ

- सटीकता (Precision) कम हो सकती है
- अप्रासंगिक परिणामों की संभावना



प्री-कोऑर्डिनेट एवं पोस्ट-कोऑर्डिनेट अनुक्रमण: तुलनात्मक दृष्टि

आधार	प्री-कोऑर्डिनेट	पोस्ट-कोऑर्डिनेट
संयोजन का चरण	अनुक्रमण के समय	खोज के समय
नियंत्रण	अधिक	कम
Precision	अधिक	अपेक्षाकृत कम
Recall	सीमित	अधिक
उपयुक्तता	पारंपरिक पुस्तकालय	डिजिटल/ऑनलाइन सिस्टम

4.6 अनुक्रमण तकनीकें (Indexing Techniques)

अनुक्रमण तकनीकें वे विधियाँ हैं जिनके माध्यम से अनुक्रमण प्रणाली को व्यवहार में लागू किया जाता है। प्रमुख अनुक्रमण तकनीकें निम्नलिखित हैं—

4.6.1 चेन अनुक्रमण (Chain Indexing)

अवधारणा

चेन अनुक्रमण डॉ. **एस. आर. रंगनाथन** द्वारा विकसित एक महत्वपूर्ण तकनीक है। यह तकनीक **वर्गीकरण योजना** पर आधारित होती है और विषयों को **सामान्य से विशिष्ट** क्रम में श्रृंखला (Chain) के रूप में प्रस्तुत करती है।

उदाहरण

यदि वर्गीकरण के अनुसार विषय है—
Library → University Library → Reference Services

तो चेन अनुक्रमण में इन सभी स्तरों को अनुक्रमण पद के रूप में प्रस्तुत किया जाएगा।

महत्त्व

- वर्गीकरण और अनुक्रमण के बीच सेतु
- विषय संरचना की स्पष्टता
- शैक्षणिक एवं अनुसंधान पुस्तकालयों में उपयोगी

4.6.2 की-वर्ड / की-फ्रेज अनुक्रमण (Keyword / Keyphrase Indexing)

अवधारणा

इस तकनीक में दस्तावेज़ में प्रयुक्त महत्वपूर्ण शब्दों या वाक्यांशों को अनुक्रमण पद के रूप में चुना जाता है। यह तकनीक **प्राकृतिक भाषा** के अधिक निकट होती है।

उदाहरण

दस्तावेज़: *Open Access Publishing in India*
अनुक्रमण पद:

- Open Access
- Scholarly Publishing
- India

विशेषताएँ

- सरल और लचीली तकनीक
- डिजिटल और वेब संसाधनों में व्यापक प्रयोग
- उपयोगकर्ता-केंद्रित खोज

4.6.3 सिटेशन अनुक्रमण (Citation Indexing)

अवधारणा

सिटेशन अनुक्रमण वह तकनीक है जिसमें दस्तावेज़ों के बीच **उद्धरण (Citation)** संबंधों के आधार पर अनुक्रमण किया जाता है। इसमें यह देखा जाता है कि कौन-सा दस्तावेज़ किस अन्य दस्तावेज़ को उद्धृत करता है।

उदाहरण

यदि कोई शोध लेख किसी अन्य लेख को संदर्भित करता है, तो दोनों के बीच एक उद्धरण संबंध स्थापित हो जाता है।

प्रमुख उदाहरण:

- Web of Science
- Scopus

महत्त्व

- शोध प्रभाव (Research Impact) का आकलन
- ज्ञान प्रवाह और विषय विकास का अध्ययन
- शोध प्रवृत्तियों की पहचान

4.7 Abstracting & Indexing सेवाएँ (A&I Services)

Abstracting & Indexing सेवाएँ वे संगठित सेवाएँ हैं जो शोध साहित्य को **संक्षेप (Abstract)** और **अनुक्रमण पदों** के माध्यम से व्यवस्थित करती हैं।

भूमिका

- शोधकर्ताओं को प्रासंगिक साहित्य की पहचान में सहायता
- सूचना पुनःप्राप्ति को सरल बनाना
- अकादमिक संचार को सुदृढ़ करना

उदाहरण

- LISA (Library and Information Science Abstracts)
- ERIC
- Chemical Abstracts

4.8 स्वचालित/मशीनी अनुक्रमण (Automatic / Machine Indexing)

अवधारणा

स्वचालित या मशीनी अनुक्रमण वह प्रक्रिया है जिसमें कंप्यूटर प्रोग्राम, एल्गोरिदम और भाषायी तकनीकों का प्रयोग कर दस्तावेज़ों का अनुक्रमण किया जाता है। इसमें मानवीय हस्तक्षेप न्यूनतम होता है और अनुक्रमण प्रक्रिया स्वचालित रूप से सम्पन्न होती है।

आधार

- सांख्यिकीय विश्लेषण
- शब्द-आवृत्ति (Term Frequency)
- Natural Language Processing (NLP)
- Machine Learning तकनीकें

लाभ

- बड़े पैमाने पर सूचना संसाधनों का शीघ्र अनुक्रमण
- समय और लागत की बचत
- डिजिटल संग्रहों के लिए उपयुक्त

सीमाएँ

- संदर्भ (Context) को समझने में कठिनाई
- बहुअर्थी शब्दों की समस्या
- मानवीय बौद्धिक विवेक का अभाव

4.9 मेटाडाटा आधारित अनुक्रमण (Metadata-based Indexing)

अवधारणा

मेटाडाटा आधारित अनुक्रमण में डिजिटल संसाधनों के वर्णनात्मक तत्वों (Metadata Elements) का उपयोग कर अनुक्रमण किया जाता है। यह डिजिटल लाइब्रेरी, संस्थागत रिपॉजिटरी और ऑनलाइन संग्रहों की आधारशिला है।

प्रमुख मेटाडाटा मानक

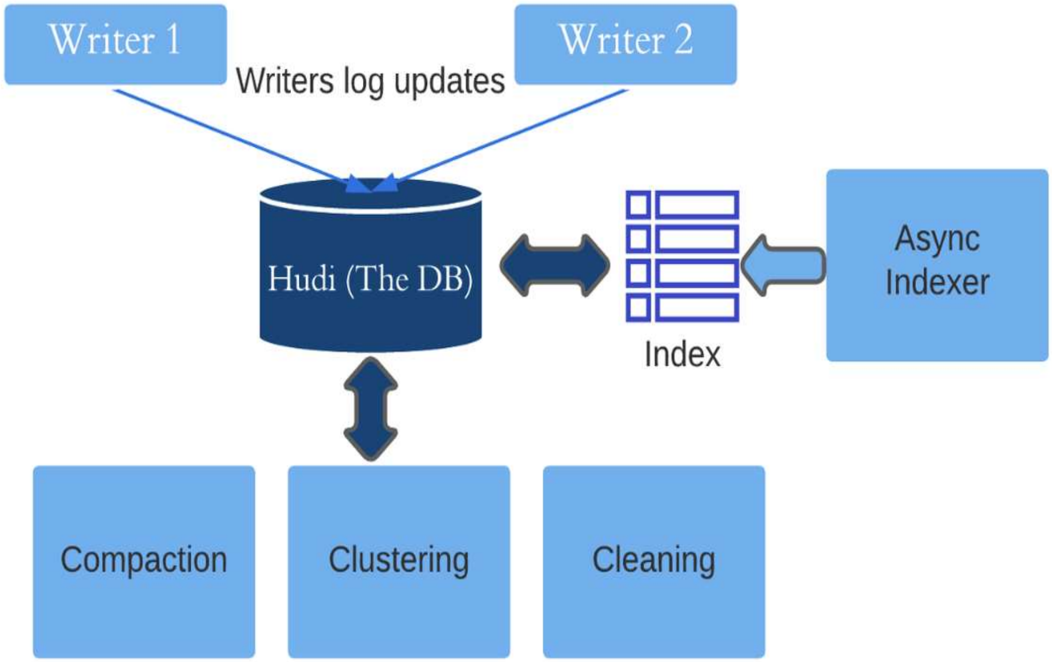
- Dublin Core
- MARC
- MODS

उदाहरण

यदि किसी ई-थीसिस का विषय “Open Access Repositories” है, तो मेटाडाटा तत्व होंगे—
Title, Subject, Creator, Date, Format

महत्त्व

- संसाधनों की खोज और पहचान में सरलता
- डिजिटल संग्रहों का मानकीकरण
- इंटरऑपरेबिलिटी में सहायता



चित्र **4.9.1:** मेटाडाटा आधारित अनुक्रमण की अवधारणा

Element	Definition
Title	A name given to a resource
Creator	An entity primarily responsible for making the content of a resource
Subject	A topic of the content of a resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of a resource
Date	A data of an event in the lifecycle of a resource
Type	The nature or genre of the content of a resource
Format	The physical or digital format of a resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to an another resource from which a resource is derived
Language	A language of the content of a resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of a resource
Rights	Information about rights held in and over a re-source

4.10 वेब अनुक्रमण एवं सर्च इंजन इंडेक्स (Web Indexing & Search Engines)

अवधारणा

वेब अनुक्रमण वह प्रक्रिया है जिसके अंतर्गत सर्च इंजन वेब पृष्ठों को खोजयोग्य बनाते हैं। इसमें तीन प्रमुख चरण होते हैं—
Crawling, Indexing और Ranking

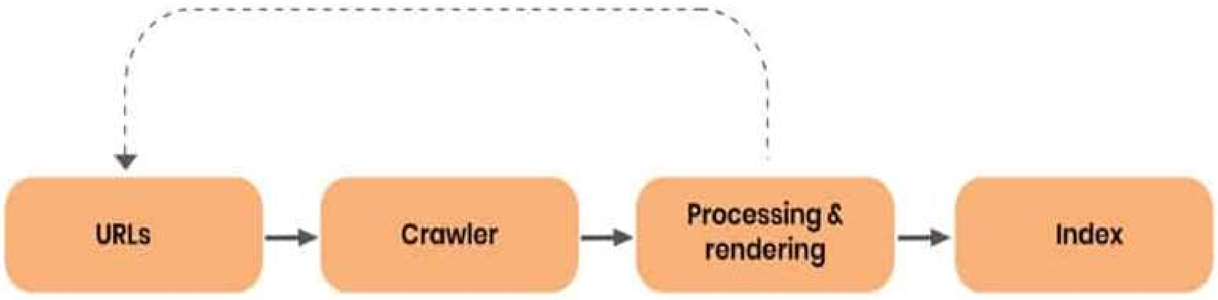
प्रक्रिया

1. Crawling – वेब पृष्ठों की पहचान
2. Indexing – विषय-वस्तु का विश्लेषण और संग्रह

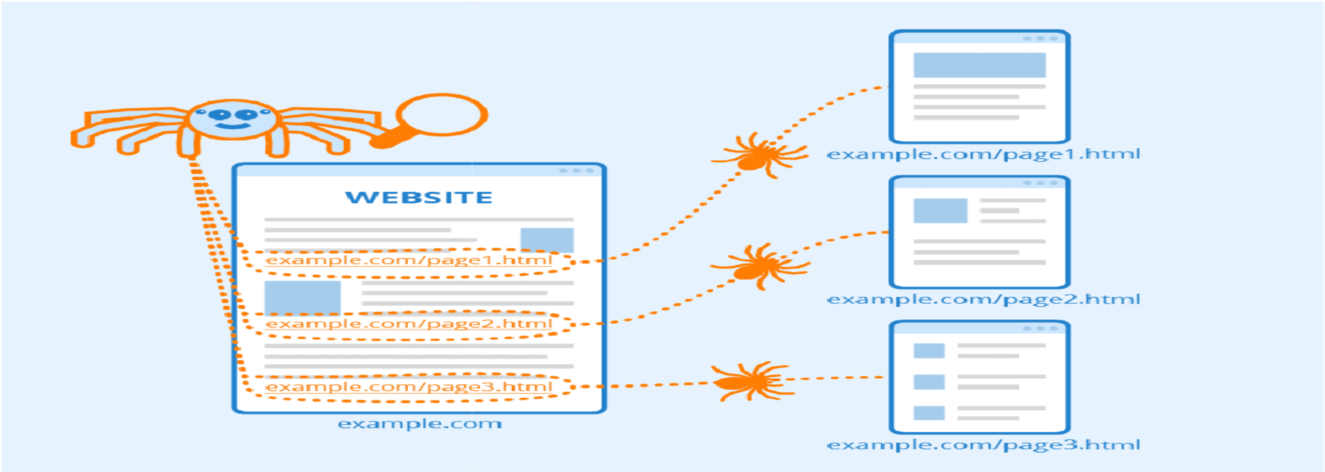
3. Ranking – प्रासंगिकता के आधार पर परिणामों का क्रम

महत्त्व

- वैश्विक सूचना तक त्वरित पहुँच
- डिजिटल शिक्षा और शोध में सहायक
- ODL शिक्षार्थियों के लिए अत्यंत उपयोगी



How Google Builts its Search Index



4.11 गुणवत्ता नियंत्रण (Quality Control in Indexing)

अनुक्रमण की गुणवत्ता सुनिश्चित करने के लिए निम्नलिखित मानदंड आवश्यक हैं—

1. Consistency (एकरूपता)

समान अवधारणाओं के लिए समान अनुक्रमण पदों का प्रयोग।

2. Exhaustivity (पर्याप्तता)

दस्तावेज़ की सभी महत्वपूर्ण अवधारणाओं का समुचित प्रतिनिधित्व।

3. Specificity (विशिष्टता)

अत्यधिक सामान्य या अत्यधिक विशिष्ट पदों से बचाव।

महत्त्व

गुणवत्ता नियंत्रण से सूचना पुनःप्राप्ति की सटीकता और उपयोगिता बढ़ती है।

4.12 व्यावहारिक उदाहरण: एक लेख का अनुक्रमण

लेख का शीर्षक:
Digital Repositories in Indian Universities

विषय विश्लेषण

मुख्य अवधारणाएँ—

- Digital Repositories
- Universities
- India

अनुक्रमण पद

Digital Repositories; Universities; India; Open Access

यह उदाहरण दर्शाता है कि किस प्रकार अनुक्रमण प्रणाली दस्तावेज़ को खोजयोग्य बनाती है।

4.13 आधुनिक प्रवृत्तियाँ (Emerging Trends)

आधुनिक अनुक्रमण प्रणाली में निम्न प्रवृत्तियाँ प्रमुख हैं—

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Semantic Indexing
- Ontology-based Indexing

ये तकनीकें अनुक्रमण को **अर्थ-आधारित (meaning-based)** और अधिक बुद्धिमान बना रही हैं।

4.14 अनुक्रमण में सामान्य त्रुटियाँ एवं सुधार

सामान्य त्रुटियाँ

- अस्पष्ट अनुक्रमण पदों का चयन
- असंगत शब्दावली
- अद्यतन की कमी

सुधार

- नियंत्रित शब्दावली का प्रयोग
- नियमित प्रशिक्षण
- गुणवत्ता नियंत्रण उपाय

4.15 सारांश

इस इकाई में अनुक्रमण प्रणाली और तकनीकों का **संपूर्ण, संतुलित और व्यावहारिक अध्ययन** प्रस्तुत किया गया है। इसमें पारंपरिक, डिजिटल और वेब आधारित अनुक्रमण प्रणालियों के साथ-साथ आधुनिक तकनीकों की भी चर्चा की गई है। यह इकाई MLIS शिक्षार्थियों को सूचना संगठन और पुनःप्राप्ति की गहन समझ प्रदान करती है।

4.16 शब्दावली (Glossary)

- **अनुक्रमण प्रणाली:** सूचना को खोजयोग्य बनाने की व्यवस्था
- **प्री-कोऑर्डिनेट अनुक्रमण:** पूर्व संयोजित अनुक्रमण
- **पोस्ट-कोऑर्डिनेट अनुक्रमण:** खोज के समय संयोजन
- **मेटाडाटा:** डेटा के बारे में डेटा
- **वेब अनुक्रमण:** वेब पृष्ठों का अनुक्रमण

4.17 निबंधात्मक प्रश्न (Essay Questions)

1. अनुक्रमण प्रणाली की अवधारणा और विकास का वर्णन कीजिए।
2. प्री-कोऑर्डिनेट और पोस्ट-कोऑर्डिनेट अनुक्रमण की तुलना कीजिए।
3. चेन अनुक्रमण की अवधारणा और महत्व स्पष्ट कीजिए।
4. स्वचालित अनुक्रमण की संभावनाएँ और सीमाएँ लिखिए।
5. वेब अनुक्रमण की प्रक्रिया का वर्णन कीजिए।

4.18 लघु प्रश्न (Short Questions)

1. अनुक्रमण प्रणाली क्या है?
2. मेटाडाटा आधारित अनुक्रमण क्या है?
3. सिटेशन अनुक्रमण का महत्व लिखिए।
4. Consistency से क्या तात्पर्य है?
5. Web Indexing क्या है?

4.19 बहुविकल्पीय प्रश्न (MCQs – 5)

1. Chain Indexing किससे संबंधित है?
   (a) वर्गीकरण ✔
   (b) वेब अनुक्रमण
   (c) मेटाडाटा
   (d) सर्च इंजन

2. Dublin Core क्या है?
   (a) वर्गीकरण योजना
   (b) मेटाडाटा मानक ✔
   (c) सर्च इंजन
   (d) अनुक्रमण सेवा

3. सिटेशन अनुक्रमण का प्रयोग किया जाता है—
   (a) पुस्तक उधार में
   (b) शोध प्रभाव अध्ययन में ✔
   (c) वर्गीकरण में
   (d) सूचीकरण में

4. Web Indexing का पहला चरण है—
   (a) Ranking
   (b) Crawling ✔
   (c) Storage
   (d) Filtering

5. Exhaustivity का संबंध है—
   (a) सटीकता से
   (b) पर्याप्त विषय कवरेज से ✔
   (c) शब्द चयन से
   (d) भाषा से

4.20 संदर्भ ग्रंथ सूची (References)

1. **Lancaster, F. W.** (2003). *Indexing and Abstracting in Theory and Practice*. London: Facet Publishing.
2. **Chowdhury, G. G.** (2010). *Introduction to Modern Information Retrieval*. London: Facet Publishing.
3. **Aitchison, J., Gilchrist, A., & Bawden, D.** (2000). *Thesaurus Construction and Use: A Practical Manual*. London: Aslib.
4. **Taylor, A. G., & Joudrey, D. N.** (2009). *The Organization of Information*. Westport: Libraries Unlimited.
5. **Rowley, J.** (1994). *Organizing Knowledge: An Introduction to Managing Access to Information*. Aldershot: Ashgate.
6. **Baeza-Yates, R., & Ribeiro-Neto, B.** (2011). *Modern Information Retrieval*. Boston: Addison-Wesley.
7. **Salton, G., & McGill, M. J.** (1983). *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.
8. **Garfield, E.** (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley.
9. **रंगनाथन, एस. आर.** – *पुस्तकालय विज्ञान के पाँच नियम*। नई दिल्ली: एस. एस. पब्लिशर्स।
10. **पी. एस. जी. कुमार** – *सूचना संगठन एवं पुनःप्राप्ति (Information Organization & Retrieval)*। नई दिल्ली: बी. आर. पब्लिशिंग।
11. **आर. एन. मिश्रा** – *ज्ञान संगठन: सिद्धांत एवं व्यवहार*। नई दिल्ली: अटलांटिक पब्लिशर्स।
12. **के. के. वर्मा** – *पुस्तकालय अनुक्रमण एवं वर्गीकरण*। दिल्ली: राज पब्लिकेशन्स।
13. **जे. पी. शर्मा** – *पुस्तकालय एवं सूचना विज्ञान के मूल सिद्धांत*। जयपुर: पंचशील प्रकाशन।
14. **एस. डी. शर्मा** – *आधुनिक पुस्तकालय एवं सूचना सेवाएँ*। नई दिल्ली: एसोसिएटेड पब्लिशर्स।

इकाई–5 : अनुक्रमण प्रणाली का मूल्यांकन (Evaluation of Indexing / IR Support)

इकाई की रूपरेखा

5.1 प्रस्तावना
5.2 उद्देश्य
5.3 मूल्यांकन: अर्थ एवं आवश्यकता **(Why Evaluation?)**
5.4 मूल्यांकन के उद्देश्य एवं परिप्रेक्ष्य
5.5 मूल्यांकन के प्रमुख मानदंड **(Criteria)**
5.6 Recall एवं **Precision:** अवधारणा एवं महत्व
5.7 Relevance, Noise, Silence: व्याख्या
5.8 उपयोगकर्ता-आधारित मूल्यांकन **(User-centered Evaluation)**
5.9 प्रणाली-आधारित मूल्यांकन **(System-centered Evaluation)**
5.10 गुणवत्ता संकेतक: **Consistency, Coverage, Timeliness**
5.11 परीक्षण सेट/टेस्ट कलेक्शन **(Test Collection)** की भूमिका
5.12 IR systems में अनुक्रमण गुणवत्ता का प्रभाव
5.13 मूल्यांकन की सीमाएँ एवं व्यावहारिक चुनौतियाँ
5.14 केस/उदाहरण: किसी **indexing system** का मूल्यांकन कैसे करें
5.15 सारांश
5.16 शब्दावली
5.17 निबंधात्मक प्रश्न एवं उत्तर
5.18 लघु प्रश्न **(Short Questions)**
5.19 बहुविकल्पीय प्रश्न **(MCQs)**
5.20 संदर्भ ग्रंथ सूची

5.1 प्रस्तावना

अनुक्रमण प्रणाली (Indexing System) का प्रमुख उद्देश्य सूचना संसाधनों को इस प्रकार संगठित करना है कि उपयोगकर्ता उन्हें **शीघ्र, सटीक और प्रासंगिक** रूप में प्राप्त कर सके। किंतु केवल अनुक्रमण का निर्माण ही पर्याप्त नहीं है; यह भी आवश्यक है कि यह प्रणाली **वास्तव में कितनी प्रभावी है**, इसका नियमित **मूल्यांकन (Evaluation)** किया जाए।

सूचना संसाधनों की तीव्र वृद्धि, डिजिटल संग्रहों का विस्तार तथा ऑनलाइन सूचना पुनःप्राप्ति प्रणालियों (IR Systems) के बढ़ते उपयोग ने अनुक्रमण प्रणाली के मूल्यांकन को और अधिक महत्वपूर्ण बना दिया है।
बिना मूल्यांकन के कोई भी अनुक्रमण प्रणाली दीर्घकाल में उपयोगी सिद्ध नहीं हो सकती।

5.2 उद्देश्य

इस इकाई के अध्ययन के पश्चात शिक्षार्थी सक्षम होंगे—

- अनुक्रमण प्रणाली के मूल्यांकन की अवधारणा को समझने में
- मूल्यांकन की आवश्यकता एवं उद्देश्यों को स्पष्ट करने में
- Recall एवं Precision जैसे प्रमुख मापदंडों का विश्लेषण करने में
- उपयोगकर्ता-आधारित एवं प्रणाली-आधारित मूल्यांकन में अंतर समझने में
- किसी अनुक्रमण/IR प्रणाली के व्यावहारिक मूल्यांकन में

5.3 मूल्यांकन: अर्थ एवं आवश्यकता (Why Evaluation?)

अर्थ (Meaning of Evaluation)

मूल्यांकन वह व्यवस्थित प्रक्रिया है जिसके माध्यम से यह आकलन किया जाता है कि कोई अनुक्रमण प्रणाली या सूचना पुनःप्राप्ति प्रणाली **अपने निर्धारित उद्देश्यों को किस सीमा तक पूरा कर रही है।**

सरल शब्दों में—

“Evaluation is the measurement of effectiveness and efficiency of an indexing system.”

मूल्यांकन की आवश्यकता

अनुक्रमण प्रणाली के मूल्यांकन की आवश्यकता निम्न कारणों से होती है—

1. यह जानने के लिए कि प्रणाली उपयोगकर्ता की सूचना आवश्यकताओं को कितनी अच्छी तरह पूरा कर रही है
2. अनुक्रमण गुणवत्ता (Indexing Quality) की जाँच के लिए
3. Recall और Precision के स्तर का निर्धारण करने हेतु
4. प्रणाली में सुधार एवं अद्यतन के लिए
5. विभिन्न अनुक्रमण प्रणालियों की तुलना करने हेतु

5.4 मूल्यांकन के उद्देश्य एवं परिप्रेक्ष्य

(A) प्रमुख उद्देश्य

- सूचना पुनःप्राप्ति की प्रभावशीलता का आकलन
- अनुक्रमण नीति (Indexing Policy) की उपयुक्तता की जाँच
- उपयोगकर्ता संतुष्टि (User Satisfaction) का मूल्यांकन
- संसाधनों और लागत की उपयोगिता का विश्लेषण

(B) मूल्यांकन के परिप्रेक्ष्य

अनुक्रमण प्रणाली का मूल्यांकन मुख्यतः दो दृष्टिकोणों से किया जाता है—

1. **उपयोगकर्ता परिप्रेक्ष्य (User Perspective)**
2. **प्रणाली परिप्रेक्ष्य (System Perspective)**

इन दोनों परिप्रेक्ष्यों की चर्चा आगे के अनुभागों में विस्तार से की जाएगी।

5.5 मूल्यांकन के प्रमुख मानदंड (Evaluation Criteria)

अनुक्रमण एवं IR प्रणाली के मूल्यांकन हेतु निम्नलिखित मानदंड प्रमुख हैं—

- Recall
- Precision
- Relevance
- Noise
- Silence
- Consistency
- Coverage

- Timeliness

इन मानदंडों के माध्यम से यह तय किया जाता है कि प्रणाली **कितनी सटीक, व्यापक और विश्वसनीय** है।

5.6 Recall एवं Precision: अवधारणा एवं महत्व

(A) Recall (रिकॉल)

परिभाषा:
Recall उस अनुपात को दर्शाता है जिसमें प्रणाली द्वारा पुनःप्राप्त प्रासंगिक दस्तावेज़ों की संख्या, संग्रह में उपलब्ध कुल प्रासंगिक दस्तावेज़ों की संख्या के सापेक्ष होती है।

सूत्र:
Recall = (प्राप्त प्रासंगिक दस्तावेज़ / कुल प्रासंगिक दस्तावेज़)

महत्त्व

- शोध एवं अकादमिक खोज में अत्यंत उपयोगी
- यह दर्शाता है कि प्रणाली कितनी व्यापक खोज कर रही है

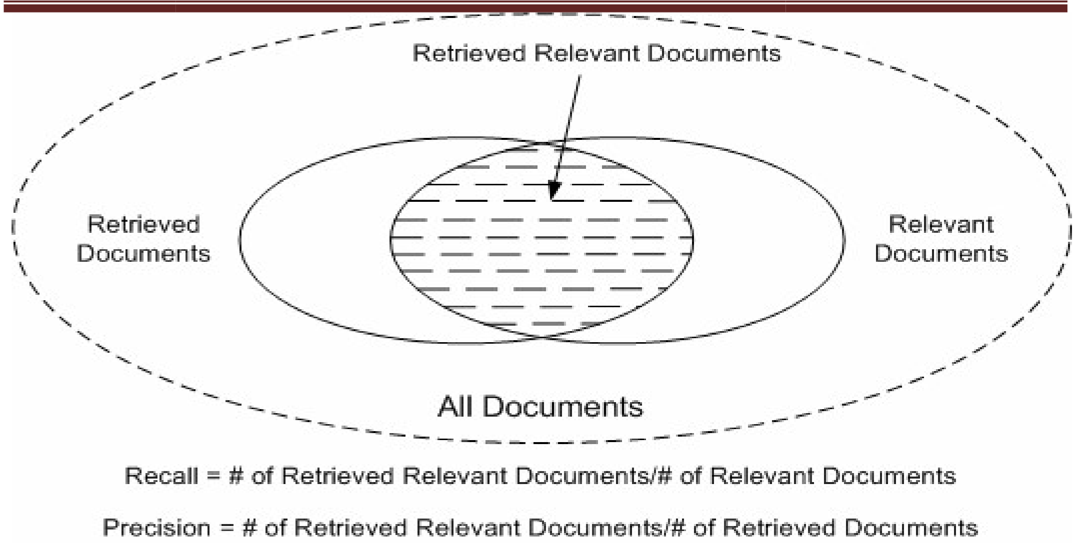
(B) Precision (प्रिसीजन)

परिभाषा:
Precision वह अनुपात है जिसमें पुनःप्राप्त दस्तावेज़ों में से वास्तव में प्रासंगिक दस्तावेज़ों की संख्या ज्ञात की जाती है।

सूत्र:
Precision = (प्राप्त प्रासंगिक दस्तावेज़ / कुल प्राप्त दस्तावेज़)

महत्त्व

- खोज परिणामों की सटीकता दर्शाता है
- उपयोगकर्ता संतुष्टि से सीधा संबंध



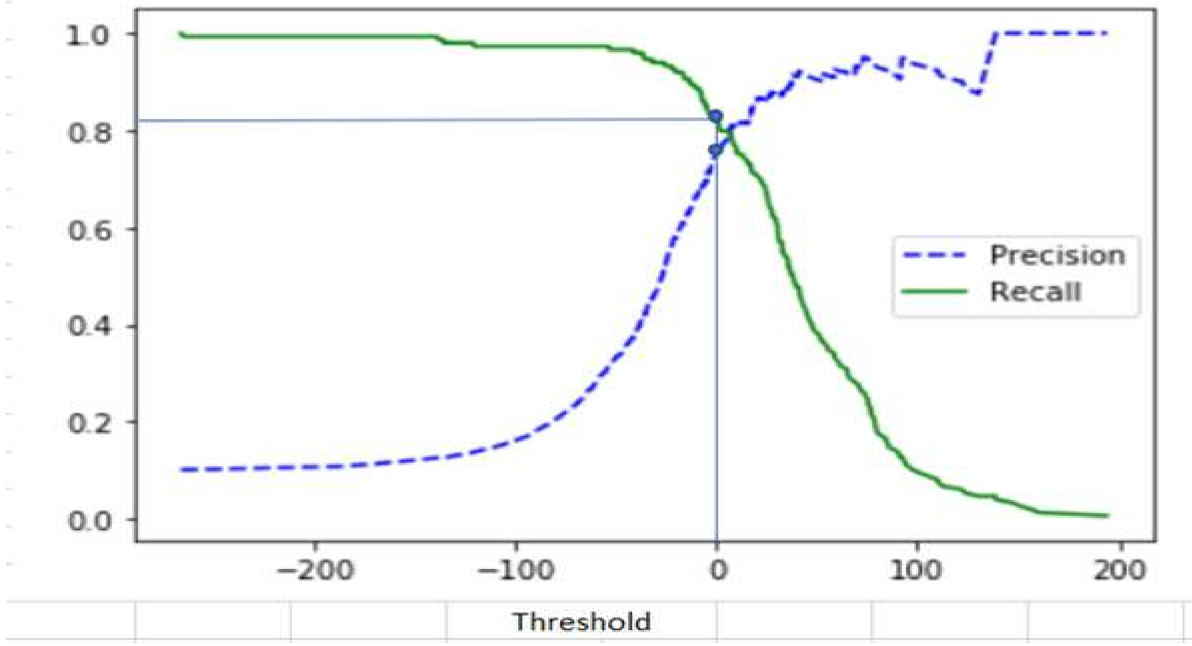


Fig.: Recall–Precision Trade-off

सामान्यतः Recall बढ़ाने पर Precision घटता है और Precision बढ़ाने पर Recall कम होता है।

5.7 Relevance, Noise एवं Silence

(A) Relevance (प्रासंगिकता)

दस्तावेज़ और उपयोगकर्ता की सूचना आवश्यकता के बीच सामंजस्य को Relevance कहा जाता है।

(B) Noise (शोर)

Noise वे दस्तावेज़ होते हैं जो पुनःप्राप्त तो होते हैं, परंतु उपयोगकर्ता के लिए प्रासंगिक नहीं होते।

(C) Silence (मौन)

Silence वे प्रासंगिक दस्तावेज़ होते हैं जो प्रणाली द्वारा पुनःप्राप्त नहीं किए जा सके।

इन दोनों स्थितियों को कम करना अनुक्रमण गुणवत्ता का प्रमुख लक्ष्य होता है।

5.8 उपयोगकर्ता-आधारित मूल्यांकन (User-Centered Evaluation)

अवधारणा

उपयोगकर्ता-आधारित मूल्यांकन वह पद्धति है जिसमें अनुक्रमण/IR प्रणाली की प्रभावशीलता को **उपयोगकर्ता की संतुष्टि, अनुभव और सफलता** के आधार पर परखा जाता है। ODL संदर्भ में यह अत्यंत महत्त्वपूर्ण है, क्योंकि शिक्षार्थी स्वयं खोज करता है और प्रणाली पर निर्भर रहता है।

प्रमुख घटक

- **User Satisfaction** (क्या उपयोगकर्ता को अपेक्षित सूचना मिली?)
- **Ease of Use** (खोज प्रक्रिया कितनी सरल है?)
- **Response Time** (परिणाम कितनी जल्दी मिले?)
- **Search Success Rate** (खोज सफल हुई या नहीं?)

मूल्यांकन की विधियाँ

- प्रश्नावली (Questionnaires)
- साक्षात्कार (Interviews)
- उपयोग लॉग विश्लेषण (Usage Logs)
- Task-based Testing (निर्धारित खोज कार्य)

उदाहरण

डिजिटल लाइब्रेरी में छात्रों से पूछा जाए—

"क्या खोज परिणाम प्रासंगिक थे?"
"क्या उपयुक्त शब्द सुझाए गए?"

निष्कर्ष: यदि अधिकांश उपयोगकर्ता संतुष्ट हैं, तो अनुक्रमण नीति प्रभावी मानी जाएगी।

5.9 प्रणाली-आधारित मूल्यांकन (System-Centered Evaluation)

अवधारणा

प्रणाली-आधारित मूल्यांकन में अनुक्रमण/IR प्रणाली को **तकनीकी और मापन योग्य मानदंडों** के आधार पर परखा जाता है—जैसे Recall, Precision, Coverage, Consistency आदि।

प्रमुख पहलू

- Indexing Policy का अनुपालन
- एल्गोरिदम की दक्षता
- संसाधन कवरेज
- Processing Speed

लाभ

- वस्तुनिष्ठ (Objective) परिणाम
- प्रणालियों की तुलना में सहायक
- तकनीकी सुधार के लिए स्पष्ट संकेत

सीमाएँ

- उपयोगकर्ता अनुभव की अनदेखी
- Contextual relevance कम आंकी जा सकती है

5.10 गुणवत्ता संकेतक (Quality Indicators)

अनुक्रमण गुणवत्ता सुनिश्चित करने के लिए निम्न संकेतक अपनाए जाते हैं—

(A) Consistency (एकरूपता)

समान अवधारणाओं के लिए समान अनुक्रमण पदों का प्रयोग।
उदाहरण: "E-learning" और "Online Learning" के लिए एक ही preferred term।

(B) Coverage (कवरेज)

संग्रह की विषयगत व्यापकता।
उदाहरण: यदि विश्वविद्यालय में विज्ञान, समाजशास्त्र और प्रबंधन—तीनों क्षेत्रों का समुचित अनुक्रमण है।

(C) Timeliness (समयानुकूलता)

नए दस्तावेज़ों का समय पर अनुक्रमण।
उदाहरण: नवीन शोध लेखों का शीघ्र इंडेक्स होना।

5.11 परीक्षण सेट / टेस्ट कलेक्शन (Test Collection) की भूमिका

अवधारणा

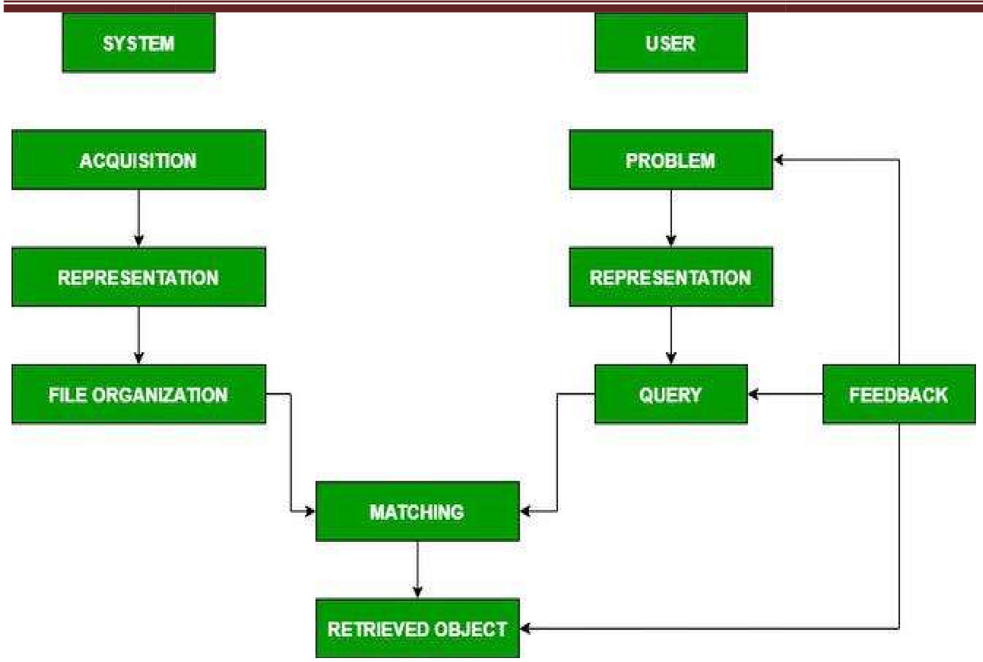
Test Collection एक **नियंत्रित डेटा-सेट** होता है, जिसका उपयोग IR प्रणाली के प्रदर्शन के आकलन के लिए किया जाता है।

मुख्य घटक

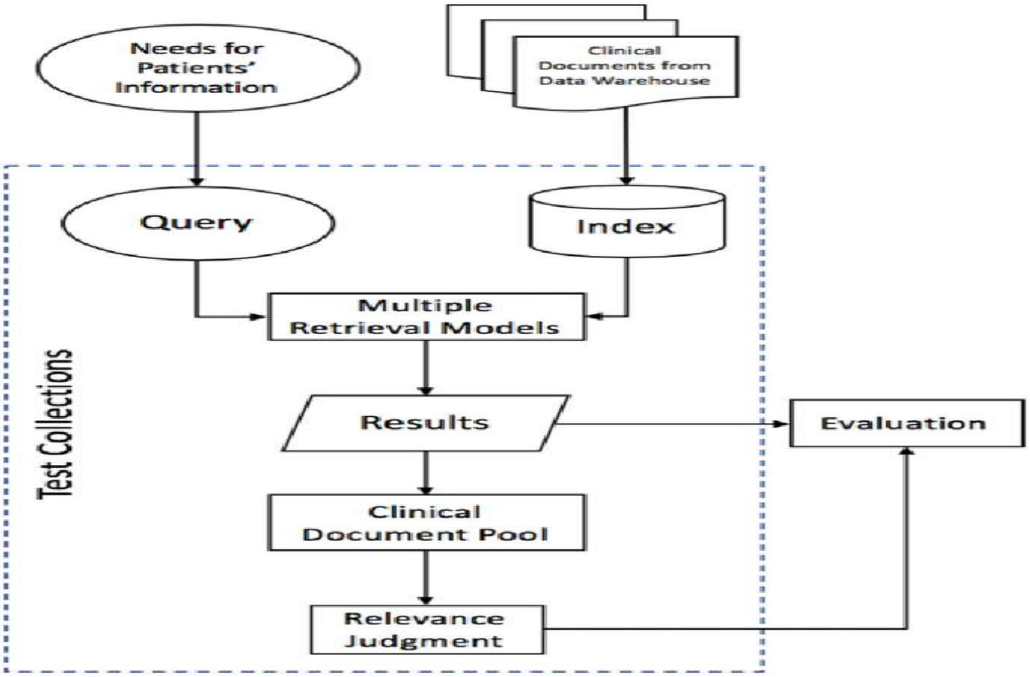
1. **Document Collection** (दस्तावेज़ों का समूह)
2. **Queries** (खोज प्रश्न)
3. **Relevance Judgements** (कौन-सा दस्तावेज़ प्रासंगिक है)

उपयोग

- Recall और Precision की गणना
- विभिन्न प्रणालियों की तुलना
- एल्गोरिदम सुधार



Fg.: IR System



Fg.: IR System Procedure

5.12 IR Systems में अनुक्रमण गुणवत्ता का प्रभाव

प्रत्यक्ष प्रभाव

- बेहतर अनुक्रमण → उच्च Precision
- अपर्याप्त अनुक्रमण → अधिक Noise

अप्रत्यक्ष प्रभाव

- उपयोगकर्ता संतुष्टि
- प्रणाली की विश्वसनीयता
- शोध की गुणवत्ता

उदाहरण

यदि अनुक्रमण पद अस्पष्ट हैं, तो उपयोगकर्ता को अप्रासंगिक परिणाम मिलेंगे—जिससे विश्वास घटेगा।

5.13 मूल्यांकन की सीमाएँ एवं व्यावहारिक चुनौतियाँ

सीमाएँ

- Relevance का व्यक्तिपरक होना
- उपयोगकर्ता आवश्यकताओं में विविधता
- संसाधन एवं समय की कमी

चुनौतियाँ

- बड़े डिजिटल संग्रह
- बहुभाषिक सामग्री
- बदलती खोज-भाषा

समाधान:
मिश्रित (Hybrid) मूल्यांकन—User + System दोनों दृष्टिकोण।

5.14 केस / उदाहरण: किसी अनुक्रमण प्रणाली का मूल्यांकन कैसे करें

इस अनुभाग में एक व्यावहारिक उदाहरण के माध्यम से यह समझाया गया है कि किसी अनुक्रमण अथवा सूचना पुनःप्राप्ति (IR) प्रणाली का मूल्यांकन वास्तविक स्थिति में कैसे किया जाता है।

केस स्टडी: विश्वविद्यालय डिजिटल लाइब्रेरी

परिस्थिति:
एक विश्वविद्यालय की डिजिटल लाइब्रेरी में शोध लेखों और थीसिस का संग्रह उपलब्ध है। उपयोगकर्ताओं (छात्रों/शोधार्थियों) की शिकायत है कि खोज परिणाम कभी-कभी अप्रासंगिक होते हैं।

मूल्यांकन के चरण

चरण 1: उद्देश्य निर्धारण

- यह जाँचना कि खोज परिणाम कितने प्रासंगिक हैं
- Recall एवं Precision का स्तर मापना
- उपयोगकर्ता संतुष्टि का आकलन करना

चरण 2: परीक्षण सेट (Test Collection) का चयन

- 200 शोध लेखों का संग्रह
- 20 सामान्य खोज प्रश्न
- विशेषज्ञों द्वारा Relevance Judgement

चरण 3: Recall एवं Precision की गणना

- Recall = 0.75
- Precision = 0.60

व्याख्या:
प्रणाली अधिकांश प्रासंगिक दस्तावेज़ पुनःप्राप्त कर रही है (Recall अच्छा है), परंतु अप्रासंगिक दस्तावेज़ भी अधिक हैं (Precision कम है)।

चरण 4: उपयोगकर्ता प्रतिक्रिया

- 65% उपयोगकर्ता संतुष्ट
- 20% आंशिक रूप से संतुष्ट

- 15% असंतुष्ट

चरण 5: निष्कर्ष एवं सुधार

- अनुक्रमण पदों को अधिक विशिष्ट बनाने की आवश्यकता
- नियंत्रित शब्दावली का सुदृढ़ प्रयोग
- Cross-references और search suggestions जोड़ना

यह उदाहरण दर्शाता है कि **मूल्यांकन केवल गणना नहीं, बल्कि सुधार का साधन है।**

5.15 सारांश (Summary)

इस इकाई में अनुक्रमण प्रणाली एवं सूचना पुनःप्राप्ति समर्थन के मूल्यांकन का सैद्धांतिक तथा व्यावहारिक अध्ययन प्रस्तुत किया गया है। Recall, Precision, Relevance, Noise एवं Silence जैसे प्रमुख मानदंडों के माध्यम से यह स्पष्ट किया गया है कि अनुक्रमण गुणवत्ता का सीधा प्रभाव सूचना पुनःप्राप्ति की सटीकता, व्यापकता एवं उपयोगकर्ता संतुष्टि पर पड़ता है।

इकाई में उपयोगकर्ता-आधारित तथा प्रणाली-आधारित मूल्यांकन दृष्टिकोणों पर प्रकाश डाला गया है और यह प्रतिपादित किया गया है कि प्रभावी मूल्यांकन के लिए दोनों दृष्टिकोणों का संतुलित प्रयोग आवश्यक है। साथ ही, Consistency, Coverage एवं Timeliness जैसे गुणवत्ता संकेतकों ने अनुक्रमण प्रणाली की विश्वसनीयता एवं प्रभावशीलता को रेखांकित किया है।

5.16 शब्दावली (Glossary)

- **Evaluation (मूल्यांकन):** प्रणाली की प्रभावशीलता का आकलन
- **Recall:** प्राप्त प्रासंगिक दस्तावेज़ों का अनुपात
- **Precision:** पुनःप्राप्त दस्तावेज़ों की सटीकता
- **Noise:** अप्रासंगिक पुनःप्राप्त दस्तावेज़
- **Silence:** अप्राप्त प्रासंगिक दस्तावेज़
- **Test Collection:** मूल्यांकन हेतु मानकीकृत डेटा-सेट

5.17 निबंधात्मक प्रश्न

1. अनुक्रमण प्रणाली के मूल्यांकन की आवश्यकता एवं महत्व स्पष्ट कीजिए।
2. Recall एवं Precision की अवधारणा उदाहरण सहित समझाइए।
3. उपयोगकर्ता-आधारित एवं प्रणाली-आधारित मूल्यांकन में अंतर कीजिए।

4. अनुक्रमण गुणवत्ता संकेतकों का वर्णन कीजिए।
5. किसी डिजिटल IR प्रणाली के मूल्यांकन की प्रक्रिया समझाइए।

5.17. 1. निबंधात्मक प्रश्नों के उत्तर

उत्तर: **1.** अनुक्रमण प्रणाली (Indexing System) का मुख्य उद्देश्य सूचना संसाधनों को इस प्रकार व्यवस्थित करना है कि उपयोगकर्ता अपनी सूचना आवश्यकता के अनुसार **सटीक, प्रासंगिक और शीघ्र सूचना** प्राप्त कर सके। परंतु केवल अनुक्रमण का निर्माण पर्याप्त नहीं होता, बल्कि यह जानना भी आवश्यक है कि अनुक्रमण प्रणाली **वास्तव में कितनी प्रभावी और उपयोगी है**। इसी उद्देश्य से अनुक्रमण प्रणाली का **मूल्यांकन (Evaluation)** किया जाता है।

अनुक्रमण प्रणाली के मूल्यांकन की आवश्यकता

अनुक्रमण प्रणाली के मूल्यांकन की आवश्यकता निम्नलिखित कारणों से होती है—

1. यह पता लगाने के लिए कि प्रणाली उपयोगकर्ता की सूचना आवश्यकताओं को कितनी अच्छी तरह पूरा कर रही है।
2. अनुक्रमण गुणवत्ता (Indexing Quality) की जाँच करने हेतु।
3. सूचना पुनःप्राप्ति में Recall एवं Precision के स्तर को मापने के लिए।
4. प्रणाली की कमियों, त्रुटियों एवं सीमाओं की पहचान करने के लिए।
5. अनुक्रमण नीति एवं शब्दावली में सुधार हेतु।
6. विभिन्न अनुक्रमण या IR प्रणालियों की तुलना करने के लिए।

अनुक्रमण प्रणाली के मूल्यांकन का महत्व

- मूल्यांकन से प्रणाली की **प्रभावशीलता (Effectiveness)** एवं **दक्षता (Efficiency)** का आकलन होता है।
- यह उपयोगकर्ता संतुष्टि बढ़ाने में सहायक होता है।
- प्रणाली को समयानुकूल एवं अद्यतन बनाए रखने में मदद करता है।
- डिजिटल एवं ऑनलाइन सूचना प्रणालियों में विश्वसनीयता बनाए रखता है।

उत्तर: 2 Recall एवं Precision सूचना पुनःप्राप्ति (Information Retrieval) प्रणाली के **दो प्रमुख मूल्यांकन मापदंड** हैं, जिनका उपयोग यह जानने के लिए किया जाता है कि प्रणाली कितनी प्रभावी ढंग से प्रासंगिक सूचना प्रदान कर रही है।

(A) Recall (रिकॉल)

परिभाषा: Recall वह अनुपात है, जो यह दर्शाता है कि संग्रह में उपलब्ध **कुल प्रासंगिक दस्तावेज़ों में से कितने दस्तावेज़ प्रणाली द्वारा पुनःप्राप्त किए गए हैं।**

सूत्र:

सूत्र (Formula):

$$\text{Recall} = \frac{\text{प्राप्त प्रासंगिक दस्तावेज़ों की संख्या}}{\text{कुल प्रासंगिक दस्तावेज़ों की संख्या}}$$

Recall = (जितने प्रासंगिक दस्तावेज़ मिले) ÷ (कुल उपलब्ध प्रासंगिक दस्तावेज़)

उदाहरण:

मान लीजिए—

- किसी डेटाबेस में **कुल 100 प्रासंगिक दस्तावेज़** उपलब्ध हैं
- खोज प्रणाली ने इनमें से **80 प्रासंगिक दस्तावेज़** पुनःप्राप्त किए

Recall = 80/100 = 0.80

Recall = 80%

इसका अर्थ है कि प्रणाली ने **80% प्रासंगिक दस्तावेज़ सफलतापूर्वक खोज लिए।**

महत्व:

- शोध एवं अकादमिक खोज में अत्यंत उपयोगी
- यह प्रणाली की **व्यापकता (Coverage)** दर्शाता है

(B) Precision (प्रिसीजन)

परिभाषा:

Precision वह अनुपात है, जो यह दर्शाता है कि **पुनःप्राप्त दस्तावेज़ों में से कितने वास्तव में प्रासंगिक हैं।**

सूत्र (Formula):

$$\text{Precision} = \frac{\text{प्राप्त प्रासंगिक दस्तावेज़ों की संख्या}}{\text{कुल प्राप्त दस्तावेज़ों की संख्या}}$$

सरल शब्दों में:

Precision = **(जितने प्रासंगिक दस्तावेज़ मिले) ÷ (कुल जितने दस्तावेज़ प्राप्त हुए)**

उदाहरण:

यदि प्रणाली ने कुल 100 दस्तावेज़ पुनःप्राप्त किए, जिनमें से 60 प्रासंगिक हैं, तो—

Precision = 60/100 = 0.60

या
Precision = 60%

अर्थ (Meaning):
इसका अर्थ है कि **प्राप्त परिणामों में से 60% दस्तावेज़ वास्तव में उपयोगी (प्रासंगिक) थे।**

Recall–Precision संबंध:

सामान्यतः Recall बढ़ाने पर Precision घटता है और Precision बढ़ाने पर Recall कम होता है।

उत्तर: 3 अनुक्रमण अथवा सूचना पुनःप्राप्ति प्रणाली का मूल्यांकन दो प्रमुख दृष्टिकोणों से किया जाता है—
उपयोगकर्ता-आधारित तथा **प्रणाली-आधारित मूल्यांकन।**

आधार	उपयोगकर्ता-आधारित मूल्यांकन	प्रणाली-आधारित मूल्यांकन
दृष्टिकोण	उपयोगकर्ता अनुभव पर आधारित	तकनीकी मापदंडों पर आधारित
केंद्र	उपयोगकर्ता संतुष्टि	प्रणाली की दक्षता
प्रमुख मापदंड	Ease of Use, Satisfaction,	Recall, Precision,

	Response Time	Coverage
प्रकृति	व्यक्तिपरक (Subjective)	वस्तुनिष्ठ (Objective)
उपयोग	सेवा सुधार	तकनीकी सुधार

उत्तर: 4 अनुक्रमण गुणवत्ता संकेतक (Indexing Quality Indicators) वे मापदंड हैं जिनके माध्यम से अनुक्रमण प्रणाली की गुणवत्ता का आकलन किया जाता है।

प्रमुख गुणवत्ता संकेतक:

(A) Consistency (एकरूपता)

समान विषय या अवधारणा के लिए समान अनुक्रमण पदों का प्रयोग।
उदाहरण: "Online Learning" और "E-learning" के लिए एक ही preferred term।

(B) Coverage (कवरेज)

संग्रह के विषय क्षेत्रों को कितनी व्यापकता से अनुक्रमित किया गया है।

(C) Timeliness (समयानुकूलता)

नए दस्तावेज़ों का समय पर अनुक्रमण।

(D) Relevance (प्रासंगिकता)

अनुक्रमण पदों और उपयोगकर्ता आवश्यकता के बीच सामंजस्य।

(E) Accuracy (सटीकता)

अनुक्रमण में त्रुटियों की न्यूनतम संभावना।
ये संकेतक अनुक्रमण प्रणाली की प्रभावशीलता सुनिश्चित करने में सहायक होते हैं।

उत्तर: 5 डिजिटल सूचना पुनःप्राप्ति (IR) प्रणाली का मूल्यांकन एक **व्यवस्थित एवं चरणबद्ध प्रक्रिया** है।

मूल्यांकन की प्रक्रिया:

चरण 1: उद्देश्य निर्धारण

- Recall एवं Precision का आकलन
- उपयोगकर्ता संतुष्टि जाँचना

चरण 2: परीक्षण सेट (Test Collection) का चयन

- दस्तावेज़ों का संग्रह
- खोज प्रश्न (Queries)
- Relevance Judgements

चरण 3: मापन

- Recall एवं Precision की गणना
- Noise एवं Silence का विश्लेषण

चरण 4: उपयोगकर्ता प्रतिक्रिया

- प्रश्नावली
- साक्षात्कार
- उपयोग लॉग विश्लेषण

चरण 5: निष्कर्ष एवं सुधार

- अनुक्रमण नीति में सुधार
- नियंत्रित शब्दावली का बेहतर प्रयोग
- खोज इंटरफेस में सुधार

5.18 लघु प्रश्न (Short Questions)

1. Evaluation से क्या तात्पर्य है?
2. Noise एवं Silence क्या हैं?
3. Test Collection क्या है?
4. Consistency का महत्व बताइए।
5. User Satisfaction क्यों आवश्यक है?

5.19 बहुविकल्पीय प्रश्न (MCQs)

1. Recall का संबंध है—
   (a) सटीकता से

(b) प्रासंगिक दस्तावेज़ों की व्यापकता से ✔
(c) उपयोगकर्ता संतुष्टि से
(d) गति से

2. Precision अधिक होने का अर्थ है—
(a) अधिक Noise
(b) अधिक Silence
(c) अधिक सटीक परिणाम ✔
(d) कम Recall

3. Noise का अर्थ है—
(a) अप्राप्त दस्तावेज़
(b) अप्रासंगिक पुनःप्राप्त दस्तावेज़ ✔
(c) प्रासंगिक दस्तावेज़
(d) खोज प्रश्न

4. Test Collection में शामिल नहीं होता—
(a) Queries
(b) Documents
(c) Relevance Judgement
(d) User Profile ✔

5. Timeliness का संबंध है—
(a) विषय चयन से
(b) समय पर अनुक्रमण से ✔
(c) Precision से
(d) Recall से

5.20 संदर्भ ग्रंथ सूची

1. **Lancaster, F. W.** (1998). Indexing and Abstracting in Theory and *Practice*. London: Facet Publishing.
2. **Chowdhury, G. G.** (2010). Introduction to Modern Information Retrieval. Facet Publishing.
3. **Salton, G.** (1989). Automatic Text Processing. Addison-Wesley.
4. **Baeza-Yates, R., & Ribeiro-Neto, B.** (2011). Modern Information Retrieval. Pearson.
5. **Manning, C. D., Raghavan, P., & Schütze, H.** (2008). Introduction to Information Retrieval. Cambridge University Press.
6. **Van Rijsbergen, C. J.** (1979). Information Retrieval. Butterworths.
7. **Rowley, J.** (2002). Information Retrieval. Ashgate Publishing.

8. **रंगनाथन, एस. आर.** – पुस्तकालय विज्ञान के पाँच नियम।
9. **पी. एस. जी. कुमार** – सूचना संगठन एवं पुनःप्राप्ति।
10. **आर. एन. मिश्रा** – ज्ञान संगठन एवं सूचना पुनःप्राप्ति प्रणाली।
11. **के. के. वर्मा** – पुस्तकालय एवं सूचना विज्ञान के सिद्धांत।
12. **डॉ. एस. डी. शर्मा** – आधुनिक पुस्तकालय सेवाएँ एवं सूचना प्रबंधन।
13. **डॉ. ए. के. सिंह** – डिजिटल पुस्तकालय एवं सूचना पुनःप्राप्ति।

द्वितीय खण्ड
ग्रंथ सूची विवरण
(Bibliographic Description)

इकाई–6 : ग्रंथसूची विवरण के सिद्धांत एवं विकास (Principles and Development of Bibliographic Description)

इकाई की रूपरेखा

6.1 प्रस्तावना
6.2 उद्देश्य
6.3 ग्रंथसूची विवरण (Bibliographic Description): अर्थ एवं अवधारणा
6.4 ग्रंथसूची विवरण की आवश्यकता एवं महत्व
6.5 ग्रंथसूची नियंत्रण (Bibliographic Control) की भूमिका
6.6 ग्रंथसूची विवरण के मूल सिद्धांत
  6.6.1 पहचान का सिद्धांत (Principle of Identification)
  6.6.2 एकरूपता का सिद्धांत (Principle of Uniformity)
  6.6.3 पर्याप्तता एवं प्रासंगिकता का सिद्धांत
6.7 ग्रंथसूची विवरण का ऐतिहासिक विकास
  6.7.1 प्रारंभिक सूचीकरण प्रयास
  6.7.2 मुद्रण युग और कार्ड कैटलॉग
  6.7.3 मानकीकरण की ओर विकास
6.8 अंतरराष्ट्रीय ग्रंथसूची नियंत्रण (UBC) की अवधारणा
6.9 ग्रंथसूची विवरण एवं सूचना पुनर्प्राप्ति का संबंध
6.10 आधुनिक पुस्तकालयों में ग्रंथसूची विवरण की भूमिका
6.11 डिजिटल परिवेश में ग्रंथसूची विवरण
6.12 व्यावहारिक उदाहरण: एक पुस्तक का मूल ग्रंथसूची विवरण
6.13 चुनौतियाँ एवं सीमाएँ
6.14 सारांश
6.15 शब्दावली
6.16 निबंधात्मक प्रश्न
6.17 लघु प्रश्न एवं MCQ

6.18 अभ्यास प्रश्नों के उत्तर
6.19 संदर्भ ग्रंथ सूची

6.1 प्रस्तावना

आधुनिक ज्ञान समाज में सूचना की मात्रा अत्यधिक बढ़ चुकी है। पुस्तकालयों, सूचना केंद्रों और डिजिटल रिपॉजिटरीज़ में उपलब्ध संसाधनों की पहचान, चयन और उपयोग तभी संभव है जब उनका **सुसंगठित ग्रंथसूची विवरण (Bibliographic Description)** उपलब्ध हो।

ग्रंथसूची विवरण वह आधार है जिस पर **सूचना संगठन, खोज (retrieval), संसाधन साझाकरण (resource sharing)** और **वैश्विक ग्रंथसूची नियंत्रण** की पूरी संरचना टिकी होती है।
डिजिटल परिवेश, ऑनलाइन कैटलॉग (OPAC), नेटवर्क आधारित पुस्तकालय प्रणाली और अंतरराष्ट्रीय डेटा विनिमय के युग में ग्रंथसूची विवरण की भूमिका और भी अधिक महत्वपूर्ण हो गई है

6.2 उद्देश्य

इस इकाई के अध्ययन के उपरांत शिक्षार्थी—

- ग्रंथसूची विवरण की **अवधारणा, अर्थ और दायरे** को समझ सकेंगे
- ग्रंथसूची विवरण की **आवश्यकता और महत्व** का विश्लेषण कर सकेंगे
- ग्रंथसूची नियंत्रण (Bibliographic Control) की भूमिका को स्पष्ट कर सकेंगे
- ग्रंथसूची विवरण के **मूल सिद्धांतों** को व्यावहारिक रूप से समझ सकेंगे
- ग्रंथसूची विवरण के **ऐतिहासिक विकास** का समालोचनात्मक अध्ययन कर सकेंगे
- डिजिटल एवं आधुनिक पुस्तकालय परिवेश में ग्रंथसूची विवरण की उपयोगिता को समझ सकेंगे

6.3 ग्रंथसूची विवरण : अर्थ एवं अवधारणा

अर्थ (Meaning)

ग्रंथसूची विवरण वह प्रक्रिया है जिसमें किसी दस्तावेज़ के बारे में आवश्यक पहचानात्मक (identifying) एवं वर्णनात्मक (descriptive) सूचनाएँ एक **मानकीकृत ढाँचे** में दर्ज की जाती हैं, ताकि उस दस्तावेज़ को आसानी से पहचाना, खोजा और उपयोग किया जा सके।

अवधारणा (Concept)

ग्रंथसूची विवरण केवल शीर्षक, लेखक या प्रकाशक लिख देना नहीं है, बल्कि यह—

- दस्तावेज़ की **बौद्धिक पहचान**
- उसके **भौतिक स्वरूप**
- और उसके **सूचना-संबंधी संदर्भ**

को एक संरचित रूप में प्रस्तुत करता है।
ISBD जैसे अंतरराष्ट्रीय मानकों ने इस अवधारणा को वैश्विक स्तर पर सुदृढ़ किया है ।

6.4 ग्रंथसूची विवरण की आवश्यकता एवं महत्व

ग्रंथसूची विवरण की आवश्यकता निम्न कारणों से उत्पन्न होती है—

1. **सूचना की पहचान** – समान शीर्षक/विषय वाले संसाधनों में भेद
2. **सूचना की खोज** – OPAC, डेटाबेस एवं डिजिटल लाइब्रेरी में सटीक खोज
3. **मानकीकरण** – विभिन्न पुस्तकालयों में एकरूपता
4. **संसाधन साझाकरण** – अंतर-पुस्तकालय ऋण (ILL) एवं नेटवर्किंग
5. **डिजिटल अभिगम** – ई-पुस्तकें, ई-जर्नल, वेब संसाधन

बिना ग्रंथसूची विवरण के पुस्तकालय एक **भंडार** तो बन सकता है, पर **सूचना प्रणाली** नहीं।

6.5 ग्रंथसूची नियंत्रण (Bibliographic Control) की भूमिका

ग्रंथसूची नियंत्रण (Bibliographic Control) पुस्तकालय एवं सूचना विज्ञान की वह केंद्रीय अवधारणा है, जिसके बिना **सूचना संगठन (Information Organization)** और **सूचना पुनर्प्राप्ति (Information Retrieval)** दोनों ही अधूरे रह जाते हैं।

6.5.1 ग्रंथसूची नियंत्रण का अर्थ

ग्रंथसूची नियंत्रण का अर्थ है—किसी देश/संस्था/नेटवर्क में प्रकाशित या संगृहीत दस्तावेज़ों की *पहचान, सूचीकरण, मानकीकृत वर्णन (standard description), तथा उपयोग हेतु उपलब्धता* सुनिश्चित करना।

यह “control” दो स्तरों पर समझा जा सकता है:

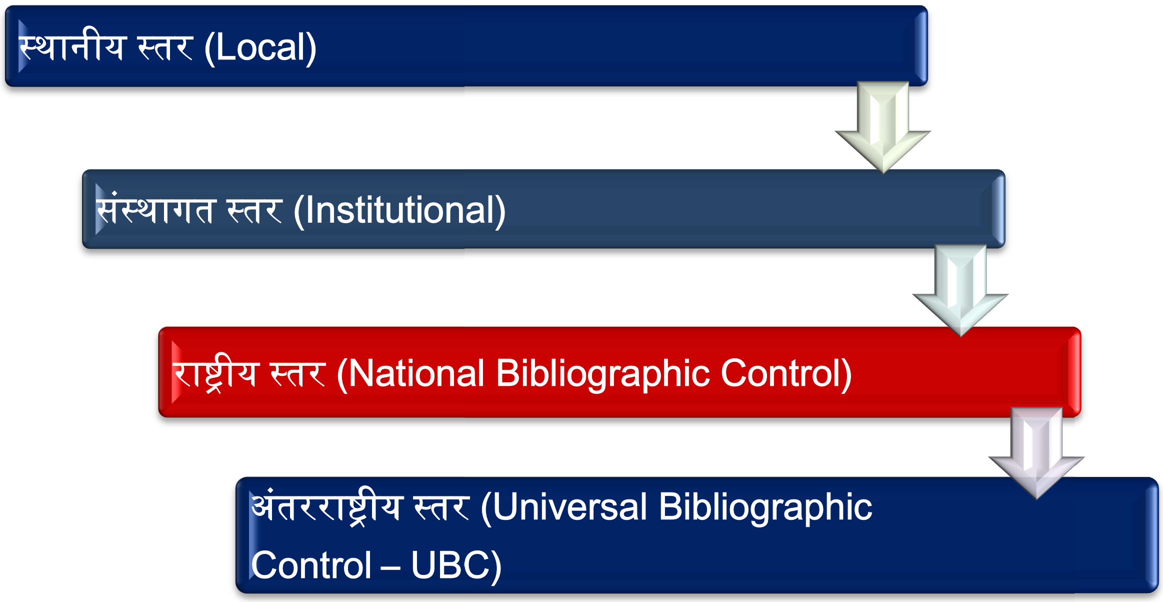
1. राष्ट्रीय स्तर (National Bibliographic Control) – राष्ट्रीय ग्रंथसूचियाँ, राष्ट्रीय पुस्तकालयों की भूमिका
2. अंतरराष्ट्रीय स्तर (Universal/International Bibliographic Control) – रिकॉर्ड्स का वैश्विक आदान-प्रदान और एकरूपता

ग्रंथसूची नियंत्रण का उद्देश्य है—

"सभी प्रकाशित दस्तावेज़ों की पहचान, अभिलेखन और उपलब्धता सुनिश्चित करना।"

यह अवधारणा **Universal Bibliographic Control (UBC)** से जुड़ी है, जिसके अंतर्गत प्रत्येक देश अपने प्रकाशनों का मानकीकृत विवरण तैयार करता है ताकि वैश्विक स्तर पर सूचना का मुक्त प्रवाह संभव हो

6.5.2 ग्रंथसूची नियंत्रण के स्तर



- स्थानीय स्तर – किसी एक पुस्तकालय का कैटलॉग
- राष्ट्रीय स्तर – National Bibliography (जैसे Indian National Bibliography)
- अंतरराष्ट्रीय स्तर – रिकॉर्ड्स का वैश्विक आदान-प्रदान (IFLA, ISBD, MARC)

Example :2 ग्रंथसूची नियंत्रण के स्तर

6.5.3 Bibliographic Control क्यों आवश्यक है?

आज सूचना "फ्लड (information flood)" जैसी स्थिति में है। इसलिए दस्तावेज़ों का **मानकीकृत विवरण** और **exchange-ready रिकॉर्ड** अत्यंत जरूरी है, ताकि:

- संसाधन खोज (Resource Discovery) सरल बने
- OPAC/डेटाबेस खोज में सटीक परिणाम मिलें
- रिकॉर्ड साझा करके लागत घटे (shared cataloguing)
- राष्ट्रीय/वैश्विक नेटवर्किंग संभव हो

ISBD भी मूलतः **universal bibliographic control** को बढ़ावा देने वाला एक प्रमुख standard है—जिसका लक्ष्य सभी देशों के प्रकाशनों के लिए *internationally acceptable* रूप में basic bibliographic data उपलब्ध कराना है।

6.5.4 Bibliographic Control का "वर्कफ़्लो" (Diagram)

उपयोगी निष्कर्ष: Bibliographic Control का अंतिम लक्ष्य केवल रिकॉर्ड बनाना नहीं, बल्कि **effective discovery + sharing + retrieval** है।

6.5.5 ग्रंथसूची नियंत्रण क्यों अनिवार्य है?

आज सूचना का स्वरूप बहुत विविध हो चुका है— पुस्तकें, ई-बुक्स, ई-जर्नल, वेबसाइट, डेटाबेस, डिजिटल रिपॉजिटरी, आदि। यदि इनका **मानकीकृत ग्रंथसूची विवरण** न हो, तो:

- उपयोगकर्ता सही दस्तावेज़ नहीं खोज पाएगा
- पुस्तकालयों में अनावश्यक दोहराव (duplication) होगा
- रिकॉर्ड साझा करना असंभव हो जाएगा
- OPAC / डिजिटल लाइब्रेरी अप्रभावी हो जाएगी

इसलिए कहा जाता है कि
“Bibliographic Control is the backbone of Library Automation.”

6.6 ग्रंथसूची विवरण के मूल सिद्धांत

ग्रंथसूची विवरण के सिद्धांत यह सुनिश्चित करते हैं कि तैयार किया गया रिकॉर्ड **सटीक, उपयोगी और सार्वभौमिक** हो।

उदाहरण:
यदि किसी पुस्तक का शीर्षक समान है, तो लेखक, संस्करण, प्रकाशन वर्ष और ISBN उसकी पहचान सुनिश्चित करते हैं।

Identification → **“ये कौन-सा दस्तावेज़ है?”**

Uniformity → **“इसी तरह का रिकॉर्ड हर जगह एक जैसा बने”**

Adequacy → **“जितना जरूरी उतना ही”**

Relevance → **“जो खोज/चयन में मदद करे वही”**

6.6.1 पहचान का सिद्धांत (Principle of Identification):

इस सिद्धांत के अनुसार, ग्रंथसूची विवरण इतना स्पष्ट होना चाहिए कि **:**

दस्तावेज़ की पहचान बिना किसी भ्रम के की जा सके।
उदाहरण: समान शीर्षक वाली पुस्तकों में संस्करण, प्रकाशन वर्ष, ISBN द्वारा पहचान।

आवश्यक तत्व:

- Title (शीर्षक)
- Author / Editor
- Edition
- Publisher & Year
- Physical description
- ISBN / ISSN

उदाहरण

यदि “Information Technology” शीर्षक की 3 पुस्तकें हैं, तो केवल शीर्षक पर्याप्त नहीं है। Edition, year, author और publisher मिलकर ही **सटीक पहचान** कराते हैं।

6.6.2 एकरूपता का सिद्धांत (Principle of Uniformity / Consistency):

एकरूपता का अर्थ है—
समान प्रकार के दस्तावेज़ों का विवरण हर **जगह एक ही क्रम, नियम और शैली** में हो। यही एकरूपता मशीन-पठनीय अभिलेख (MARC) और ISBD की आत्मा है।

यह सिद्धांत—

- रिकॉर्ड साझाकरण को आसान बनाता है
- भाषा बाधा को कम करता है
- मशीन-पठनीय अभिलेख तैयार करने में सहायक होता है

इस सिद्धांत का उद्देश्य है—

समान प्रकार के दस्तावेज़ों का विवरण हर जगह एक ही नियम, क्रम और शैली में हो।

ISBD इसी सिद्धांत पर आधारित है।
इसमें:

- तत्वों का **निश्चित क्रम (fixed order)**
- **निर्धारित विराम चिह्न (prescribed punctuation)**

दिए जाते हैं।

लाभ:

- रिकॉर्ड्स आसानी से साझा किए जा सकते हैं
- भाषा बाधा (language barrier) कम होती है
- मशीन-पठनीय रिकॉर्ड बनाना सरल होता है

6.6.3 पर्याप्तता एवं प्रासंगिकता का सिद्धांत (Principle of Adequacy & Relevance):

यह सिद्धांत संतुलन सिखाता है।

- विवरण **इतना पर्याप्त** हो कि पहचान और चयन संभव हो
- लेकिन **अनावश्यक रूप से विस्तृत** न हो

ग्रंथसूची विवरण—

- न तो अनावश्यक रूप से लंबा हो
- न ही इतना संक्षिप्त कि पहचान अधूरी रह जाए

Diagram–2 : Adequacy Balance

अल्प विवरण → पहचान में भ्रम
अत्यधिक विवरण → रिकॉर्ड भारी और जटिल
संतुलित विवरण → प्रभावी ग्रंथसूची विवरण

6.7 ग्रंथसूची विवरण का ऐतिहासिक विकास (Evolution)

ग्रंथसूची विवरण एक दिन में विकसित नहीं हुआ। यह **क्रमिक विकास** का परिणाम है।

6.7.1 प्रारंभिक सूचीकरण प्रयास

- प्रारंभिक पुस्तकालयों में सूचीकरण "स्थानीय" था(स्थानीय नियमों पर आधारित था।)—हर जगह अलग शैली
- समस्या: रिकॉर्ड साझा नहीं हो पाते, heading/description में भिन्नता
- अंतरराष्ट्रीय स्तर पर एकरूपता का अभाव था।

6.7.2 मुद्रण युग और कार्ड कैटलॉग

मुद्रण क्रांति के बाद पुस्तकों की संख्या बढ़ी। इससे आवश्यकता उत्पन्न हुई (व्यवस्थित कैटलॉगिंग)—

- Card Catalogue
- Author, Title, Subject entries
- व्यवस्थित ग्रंथसूची विवरण

कार्ड कैटलॉग के माध्यम से ग्रंथसूची विवरण व्यवस्थित हुआ, परंतु मानकीकरण की कमी बनी रही।

लेकिन समस्या बनी रही—
हर देश/संस्था के नियम अलग थे।("manual + local rules" के कारण **interchange** कठिन)

6.7.3 मानकीकरण की ओर विकास

यहीं से अंतरराष्ट्रीय मानकीकरण शुरू हुआ:

- 1961 – Paris Principles
- 1969 onwards – ISBD का विकास (1969 working group से शुरू प्रक्रिया; अलग-अलग प्रकार के संसाधनों हेतु)
- ISBD(M), ISBD(S), ISBD(NBM), ISBD(ER), ISBD(CR)
- इलेक्ट्रॉनिक संसाधनों हेतु ISBD(ER) तथा continuing resources हेतु ISBD (CR) जैसे रूप

मुख्य कारण:

- रिकॉर्ड share करने की आर्थिक आवश्यकता
- automation की जरूरत
- machine-readable records का प्रसार

उद्देश्य:
International exchange of bibliographic records

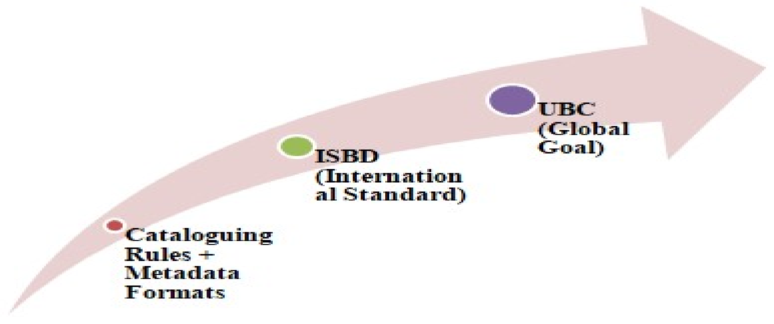
6.8 अंतरराष्ट्रीय ग्रंथसूची नियंत्रण (Universal Bibliographic Control – UBC)

UBC की मूल अवधारणा है—

हर देश अपने प्रकाशनों के definitive records बनाए और उन्हें वैश्विक स्तर पर साझा करे।

- वैश्विक स्तर पर ग्रंथसूची अभिलेखों का आदान-प्रदान
- दोहराव को कम करना
- राष्ट्रीय ग्रंथसूची एजेंसियों को सशक्त बनाना

ISBD इस अवधारणा का आधार स्तंभ है।



Fg.: ISBD और UBC का संबंध

ISBD UBC को व्यवहारिक रूप देता है।

6.8.1 UBC का मूल उद्देश्य

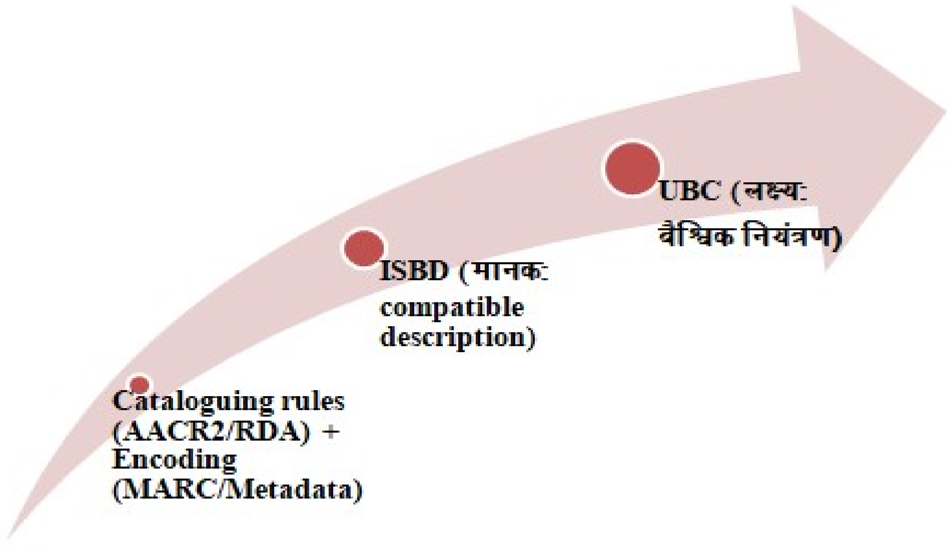
Universal Bibliographic Control (UBC) का लक्ष्य है:

- हर देश अपने प्रकाशनों के "definitive" रिकॉर्ड बनाए
- और रिकॉर्ड्स को विश्व स्तर पर साझा किया जाए ISBD को universal bibliographic control को बढ़ावा देने वाला principal standard माना गया है।

6.8.2 ISBD और UBC का संबंध

ISBD का उद्देश्य है कि:

- विभिन्न देशों/स्रोतों से आए रिकॉर्ड **interchangeable** हों
- language barrier के बावजूद समझे जा सकें
- electronic form में conversion आसान हो



6.9 ग्रंथसूची विवरण एवं सूचना पुनर्प्राप्ति (Information Retrieval) का संबंध

ग्रंथसूची विवरण और सूचना पुनर्प्राप्ति के बीच **सीधा संबंध** है।
सूचना पुनर्प्राप्ति प्रणाली (IR System) तभी प्रभावी होती है जब—

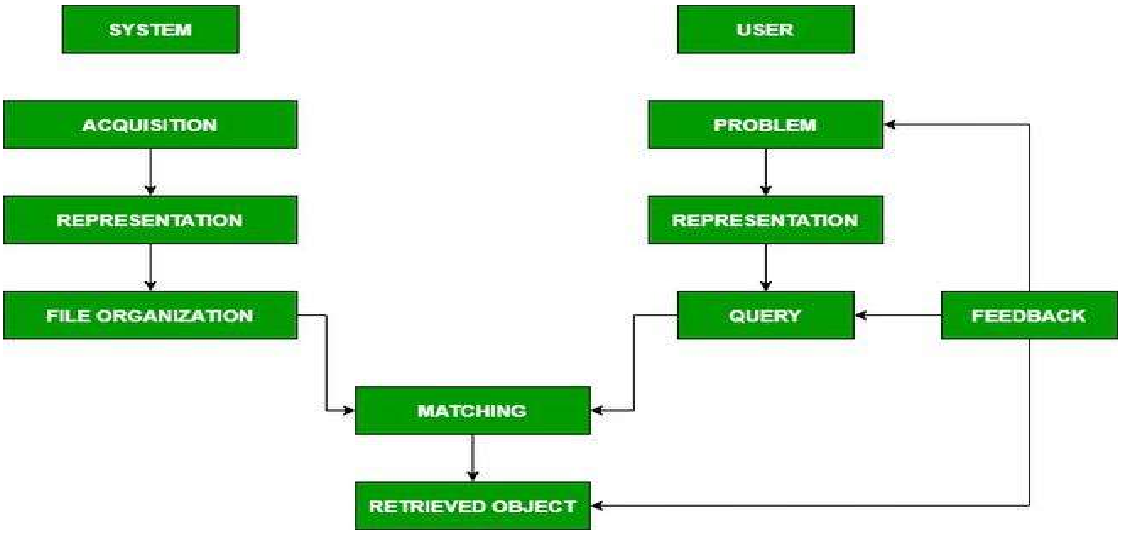
- विवरण सटीक हो
- की-वर्ड, विषय, लेखक नियंत्रित हों
- मेटाडाटा संरचित हो

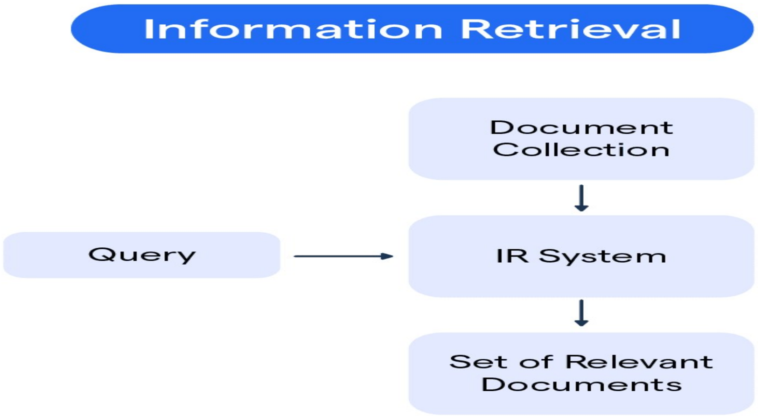
गलत या अधूरा ग्रंथसूची विवरण **खोज विफलता** का प्रमुख कारण बनता है।

Diagram: Description → Retrieval Model

यदि विवरण गलत या अधूरा है, तो सूचना पुनर्प्राप्ति स्वतः कमजोर हो जाती है।

ग्रंथसूची विवरण और सूचना पुनर्प्राप्ति का संबंध प्रत्यक्ष (direct) है।





6.9.1 संबंध की व्याख्या (क्यों "Description = Retrieval"?): सूचना पुनर्प्राप्ति में user आमतौर पर खोजता है:

- सही description → सही indexing
- सही indexing → बेहतर search results
- बेहतर search → user satisfaction

In terms of Library examples:

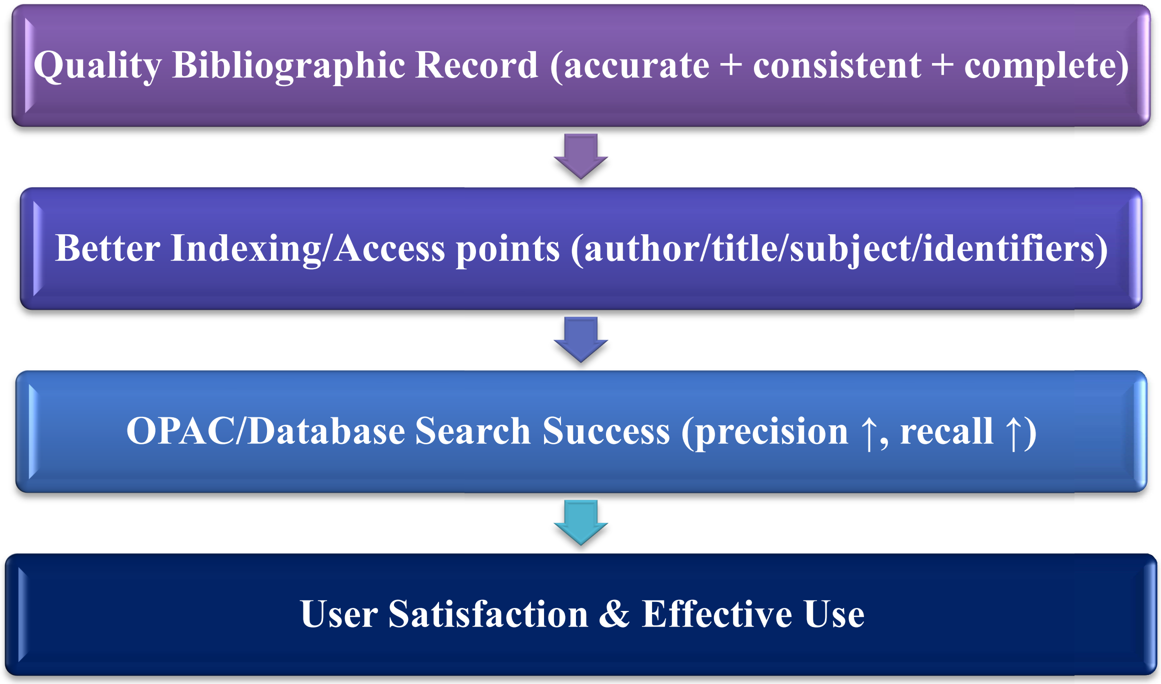
- लेखक (Author)
- शीर्षक (Title)
- विषय/की-वर्ड (Subject/Keyword)
- प्रकाशन वर्ष (Year)
- भाषा/format (Language/Media)

यदि bibliographic description में यह तत्व गलत/अधूरे हैं, तो retrieval कमजोर होगी।

Quality Record
↓
Accurate Access Points
↓
High Precision & Recall
↓
Effective Information Retrieval

यदि description गलत है, तो retrieval स्वतः कमजोर हो जाती है।

6.9.2 Retrieval-Impact मॉडल (Diagram)



6.10 आधुनिक पुस्तकालयों में ग्रंथसूची विवरण की भूमिका

आज का पुस्तकालय:

- Hybrid Library
- Digital Library
- Networked Library

यहाँ ग्रंथसूची विवरण की भूमिका—

1. OPAC आधारित खोज
2. Union Catalogue
3. Inter-Library Loan (ILL)
4. Digital Repository
5. Research Visibility

6.10 .1 आधुनिक पुस्तकालयों में ग्रंथसूची विवरण की भूमिका:

आज libraries “collection-centred” नहीं, “user-centred + network-centred” हैं। इसलिए bibliographic description के लाभ:

1. **OPAC खोज** – author/title/subject आधारित retrieval
2. **Union Catalogue** – multi-library record sharing
3. **Inter-Library Loan (ILL) & Resource Sharing** – सही item की पहचान
4. **Collection Management** (Digital Repository)– editions, duplicates, formats का नियंत्रण
5. **Research Visibility(Analytics)** – usage / holdings analysis के लिए structured data में प्रमुख भूमिका निभाता है। MARC 21 और ISBD का प्रयोग इसी संदर्भ में किया जाता है।

6.11 डिजिटल परिवेश में ग्रंथसूची विवरण

डिजिटल संसाधनों में—

- URL / DOI
- Access mode
- File format
- Version control

जैसे नए तत्व जुड़ जाते हैं। इसलिए मेटाडाटा, लिंक्ड डेटा और मानकीकृत विवरण अत्यंत आवश्यक है।

Diagram–5 : Print vs Digital Description

Print Resource:

Author → Title → Publisher → Pages

Digital Resource:

Author → Title → URL/DOI → Access Mode → Format

डिजिटल संसाधनों में 3 नई चुनौतियाँ उभरती हैं:

(A) संसाधन का स्वरूप (carrier/medium) बदलता है

ISBD विभिन्न प्रकार की published resources (electronic resources, moving images, etc.) को cover करता है

(B) Location "URL/DOI" के रूप में

इसलिए metadata में electronic access जोड़ना जरूरी हो जाता है (आगे Unit 9 में MARC 856/DC आदि)।

(C) Interoperability

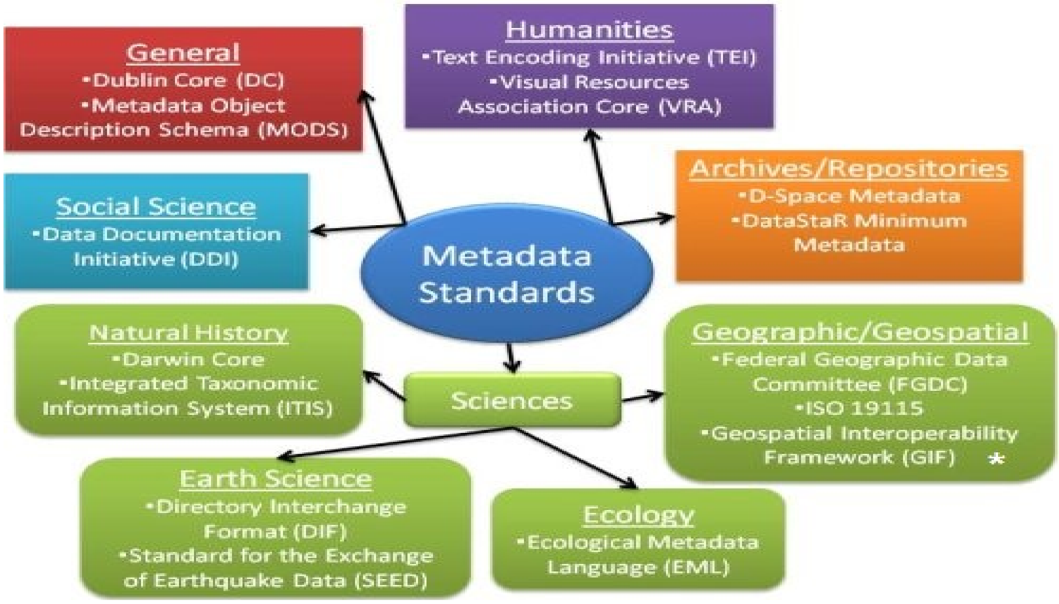
ISBD international exchange के लिए worldwide compatible descriptive cataloguing का stipulation देता है।

डिजिटल-पर्यावरण diagram:

Print era: Author/Title/Publisher/Pages

Digital era: + URL/DOI + File format + Access rights + Versioning

6.11.1 डिजिटल परिवेश में ग्रंथसूची विवरण:डिजिटल संसाधनों ने नई चुनौतियाँ दी हैं—



नई आवश्यकताएँ:

- URL / DOI
- Access mode
- File format
- Version control

इसलिए ISBD + Metadata standards अत्यंत आवश्यक हो गए हैं।

6.12 व्यावहारिक उदाहरण: एक पुस्तक का मूल ग्रंथसूची विवरण उदाहरण (ISBD क्रम में):

6.12.1 उदाहरण 1: Printed book (सरल लेकिन पूर्ण):
यह उदाहरण पहचान, एकरूपता और पर्याप्तता—तीनों सिद्धांतों को दर्शाता है।

लेखक: Ranganathan, S.R.
शीर्षक: *Colon Classification*
संस्करण: 6th ed.
प्रकाशन: Bangalore : Sarada Ranganathan Endowment, 1987.
भौतिक विवरण: xii, 420 p. ; 25 cm.
मानक संख्या: ISBN (यदि उपलब्ध)

6.12.2 उदाहरण 2: Edited book/Chapter (प्रायोगिक समझ)

यदि आपके पास "Library Standards" नामक edited volume है, तो record में editor + chapter info जोड़ना पड़ता है—यह later multilevel description के concept से जुड़ता है (advanced learning)।

6.13 चुनौतियाँ एवं सीमाएँ (Challenges & Limitations)

1. **सूचना विस्फोट** – हर प्रकार के संसाधन का विवरण कठिन
2. **बहु-मानक पर्यावरण** – AACR2, RDA, ISBD, MARC, Dublin Core साथ-साथ
3. **Language/Script issues** – transliteration, multilingual records
4. **Digital instability** – URLs बदलना, versions बदलना
5. **Human resource gap** – trained cataloguers की कमी
6. **Quality control** – consistency/exhaustivity का संतुलन

6.14 सारांश:

यह इकाई ग्रंथसूची विवरण की **सैद्धांतिक नींव**, उसके **ऐतिहासिक विकास** तथा **आधुनिक पुस्तकालय परिवेश में उसकी प्रासंगिकता** को स्पष्ट रूप से प्रस्तुत करती है। अध्ययन से यह स्पष्ट होता है कि ग्रंथसूची विवरण केवल एक तकनीकी या यांत्रिक प्रक्रिया नहीं है, बल्कि यह **सूचना संगठन और ग्रंथसूची नियंत्रण (Bibliographic Control)** का मूल दर्शन है।

इस इकाई में यह प्रतिपादित किया गया है कि ग्रंथसूची विवरण के माध्यम से सूचना संसाधनों की पहचान, चयन और खोज संभव होती है। **ISBD जैसे अंतरराष्ट्रीय मानक** ग्रंथसूची अभिलेखों को **interchangeable**, **language barrier से स्वतंत्र**, तथा **electronic environment में उपयोग योग्य** बनाते हैं, जिससे वैश्विक स्तर पर सूचना संसाधनों का आदान-प्रदान सुगम होता है।

सार रूप में, यह इकाई स्पष्ट करती है कि—

- ग्रंथसूची विवरण किसी भी सूचना प्रणाली की **आधारशिला** है
- ISBD ग्रंथसूची विवरण में **अंतरराष्ट्रीय एकरूपता** प्रदान करता है
- UBC का लक्ष्य **वैश्विक सूचना साझाकरण** को सुदृढ़ करना है
- आधुनिक एवं डिजिटल पुस्तकालयों में ग्रंथसूची विवरण का महत्व निरंतर बढ़ता जा रहा है

इस प्रकार, ग्रंथसूची विवरण को आधुनिक पुस्तकालय एवं सूचना सेवाओं का एक **अनिवार्य एवं केन्द्रीय घटक** माना जा सकता है।

6.15 शब्दावली (Glossary)

• **Digital Resource (डिजिटल संसाधन)** – इलेक्ट्रॉनिक रूप में उपलब्ध सूचना सामग्री, जैसे ई-बुक, ई-जर्नल, डेटासेट, वेबपेज।

• **Persistent Identifier (स्थायी पहचानकर्ता)** – ऐसा पहचान तंत्र जो समय के साथ नहीं बदलता, जैसे DOI।

• **DOI (Digital Object Identifier)** – डिजिटल दस्तावेज़ों की स्थायी पहचान हेतु प्रयुक्त अद्वितीय कोड।

• **URL (Uniform Resource Locator)** –इंटरनेट पर किसी संसाधन का वेब पता।

• **Access Mode (प्रवेश विधि)** – डिजिटल संसाधन तक पहुँचने का तरीका, जैसे open access, restricted access।

• **File Format (फ़ाइल प्रारूप)** –डिजिटल फ़ाइल का तकनीकी स्वरूप, जैसे PDF, HTML, XML।

• **Version Control (संस्करण नियंत्रण)** –डिजिटल संसाधन के विभिन्न संस्करणों का प्रबंधन एवं पहचान।

• **Interoperability (अंतर-प्रचालन क्षमता)** –विभिन्न प्रणालियों एवं सॉफ्टवेयर के बीच डेटा के आदान-प्रदान की क्षमता।

• **Metadata Standard (मेटाडाटा मानक)** –मेटाडाटा तत्वों के संरचनात्मक नियम, जैसे Dublin Core, MODS।

• **Dublin Core (DC)** – एक सरल एवं व्यापक रूप से प्रयुक्त मेटाडाटा मानक, जिसमें 15 मुख्य तत्व होते हैं।

• **Repository (रिपॉज़िटरी)** – डिजिटल संसाधनों को संग्रहित, संरक्षित और उपलब्ध कराने वाला तंत्र।

• **Digital Library (डिजिटल पुस्तकालय)** – डिजिटल संसाधनों का संगठित संग्रह, जिसे नेटवर्क के माध्यम से उपयोग किया जा सके।

• **Resource Discovery (संसाधन खोज)** – उपयोगकर्ता द्वारा सूचना संसाधनों की पहचान एवं खोज की प्रक्रिया।

6.16 निबंधात्मक प्रश्न

प्रश्न 1 : ग्रंथसूची नियंत्रण की अवधारणा स्पष्ट करते हुए आधुनिक पुस्तकालयों में इसकी भूमिका का विवेचन कीजिए।

उत्तर : ग्रंथसूची नियंत्रण (Bibliographic Control) से आशय पुस्तकालय में उपलब्ध सूचना संसाधनों की ऐसी व्यवस्थित पहचान, संगठन और अभिलेखन से है, जिससे उनकी **खोज, चयन और उपयोग** सरल एवं प्रभावी हो सके। यह पुस्तकालय विज्ञान की एक मूलभूत अवधारणा है, जिसका उद्देश्य सूचना के तीव्र प्रसार के युग में संसाधनों को नियंत्रित एवं सुलभ बनाना है।

पारंपरिक पुस्तकालयों में ग्रंथसूची नियंत्रण का कार्य कार्ड कैटलॉग, मुद्रित सूचियों एवं संदर्भ ग्रंथों के माध्यम से किया जाता था। परंतु आधुनिक पुस्तकालय डिजिटल, नेटवर्क-आधारित और स्वचालित हो चुके हैं। आज OPAC, डिजिटल पुस्तकालय, संस्थागत रिपॉज़िटरी और यूनियन कैटलॉग के माध्यम से ग्रंथसूची नियंत्रण का स्वरूप व्यापक और तकनीकी हो गया है।

आधुनिक पुस्तकालयों में ग्रंथसूची नियंत्रण की भूमिका निम्नलिखित है—

- संसाधनों की **सटीक पहचान (Identification)**
- एकरूप एवं मानकीकृत विवरण (Standardized description)
- **सूचना पुनर्प्राप्ति** को प्रभावी बनाना
- **Resource sharing** एवं नेटवर्किंग को संभव बनाना
- डिजिटल परिवेश में **interoperability** सुनिश्चित करना

इस प्रकार कहा जा सकता है कि आधुनिक पुस्तकालयों में ग्रंथसूची नियंत्रण केवल संगठन का माध्यम नहीं, बल्कि **ज्ञान तक पहुँच का आधार** है।

प्रश्न 2 : ISBD के उद्देश्य एवं "Universal Bibliographic Control" के साथ उसके संबंध को स्पष्ट कीजिए।

उत्तर : ISBD (International Standard Bibliographic Description) एक अंतरराष्ट्रीय मानक है, जिसे IFLA द्वारा विकसित किया गया। इसका मुख्य उद्देश्य ग्रंथसूची विवरण को इस प्रकार मानकीकृत करना है कि वह **वैश्विक स्तर पर समझने योग्य, साझा करने योग्य और मशीन-पठनीय** बन सके।

ISBD के प्रमुख उद्देश्य हैं—

- ग्रंथसूची विवरण में **एकरूपता (Uniformity)**
- विवरण तत्वों का **निश्चित क्रम एवं विराम-चिह्न**
- भाषा-बाधा को कम करना
- रिकॉर्ड की **अंतर-परिवर्तनीयता (Interchangeability)** बढ़ाना

Universal Bibliographic Control (UBC) की अवधारणा यह मानती है कि—

प्रत्येक दस्तावेज़ का ग्रंथसूची विवरण केवल एक बार बनाया जाए और उसे विश्व स्तर पर साझा किया जाए।

ISBD और UBC का संबंध अत्यंत घनिष्ठ है। ISBD, UBC की **व्यावहारिक आधारशिला** है, क्योंकि—

- ISBD के बिना रिकॉर्ड वैश्विक रूप से स्वीकार्य नहीं हो सकते
- UBC के लक्ष्य को प्राप्त करने के लिए ISBD द्वारा निर्धारित मानकीकरण आवश्यक है

अतः कहा जा सकता है कि **ISBD, Universal Bibliographic Control को व्यवहार में लागू करने का प्रमुख साधन है।**

प्रश्न 3 : ग्रंथसूची विवरण के सिद्धांतों (Identification, Uniformity, Adequacy, Relevance) की व्याख्या उदाहरण सहित कीजिए।

उत्तर : ग्रंथसूची विवरण कुछ मूलभूत सिद्धांतों पर आधारित होता है, जो रिकॉर्ड की गुणवत्ता और उपयोगिता सुनिश्चित करते हैं।

(i) Identification (पहचान)
इस सिद्धांत के अनुसार ग्रंथसूची विवरण ऐसा होना चाहिए कि संसाधन की **स्पष्ट पहचान** हो सके।
उदाहरण:
शीर्षक, लेखक, प्रकाशन विवरण से किसी पुस्तक को अन्य पुस्तकों से अलग पहचाना जा सके।
(ii) Uniformity (एकरूपता)
सभी रिकॉर्ड समान नियमों एवं मानकों के अनुसार बनाए जाएँ, ताकि उपयोगकर्ता भ्रमित न हो।
उदाहरण:
ISBD के अनुसार सभी पुस्तकों में प्रकाशन विवरण एक ही क्रम में दिया जाना।
(iii) Adequacy (पर्याप्तता)
रिकॉर्ड में न तो अनावश्यक विवरण हो और न ही जानकारी की कमी।

उदाहरण:
केवल शीर्षक लिखना अपर्याप्त है; लेखक, वर्ष, ISBN आदि भी आवश्यक हैं।

(iv) Relevance (प्रासंगिकता)
दिया गया विवरण उपयोगकर्ता की सूचना आवश्यकता से संबंधित होना चाहिए।
उदाहरण:
डिजिटल संसाधनों में URL, Access mode, File format का उल्लेख।
इन सिद्धांतों के पालन से ग्रंथसूची विवरण **विश्वसनीय, उपयोगी और प्रभावी** बनता है।

प्रश्न 4 : डिजिटल संसाधनों के संदर्भ में ग्रंथसूची विवरण की नई चुनौतियाँ एवं आवश्यक संशोधनों का विश्लेषण कीजिए।

उत्तर : डिजिटल संसाधनों के आगमन से ग्रंथसूची विवरण के क्षेत्र में अनेक नई चुनौतियाँ उत्पन्न हुई हैं। पारंपरिक मुद्रित संसाधनों की तुलना में डिजिटल संसाधन **अस्थायी, परिवर्तनीय और बहु-स्वरूपीय** होते हैं।

मुख्य चुनौतियाँ—

- URL का बदलना (Link rot)
- संसाधन के अनेक संस्करण
- भौतिक विवरण का अभाव
- Access restrictions
- दीर्घकालिक संरक्षण (Preservation)

इन चुनौतियों के कारण ग्रंथसूची विवरण में **संशोधन आवश्यक** हो गए हैं, जैसे—

- DOI जैसे **स्थायी पहचानकर्ता**
- Metadata standards (Dublin Core, MODS) का प्रयोग
- Access mode एवं file format का उल्लेख
- Version control की जानकारी

इस प्रकार डिजिटल परिवेश में **ISBD + Metadata standards** का संयुक्त प्रयोग अनिवार्य हो गया है।

प्रश्न 5 : ग्रंथसूची विवरण और सूचना पुनर्प्राप्ति के संबंध को "quality record → retrieval success" मॉडल के आधार पर समझाइए।

उत्तर : सूचना पुनर्प्राप्ति (Information Retrieval) की सफलता सीधे तौर पर ग्रंथसूची रिकॉर्ड की गुणवत्ता पर निर्भर करती है। इसे सरल रूप में **"Quality Record → Retrieval Success" मॉडल** द्वारा समझा जा सकता है।

यदि ग्रंथसूची रिकॉर्ड—

- सटीक
- पूर्ण
- मानकीकृत
- प्रासंगिक

है, तो उपयोगकर्ता को सूचना **तेज़ और सही** रूप में प्राप्त होती है।

मॉडल की व्याख्या—

- **Quality Record:** सही metadata, नियंत्रित शब्दावली, मानकीकृत विवरण
- **Effective Search:** OPAC / डेटाबेस में बेहतर matching
- **Retrieval Success:** उपयोगकर्ता की सूचना आवश्यकता की पूर्ति

उदाहरण:
यदि किसी ई-जर्नल का रिकॉर्ड सही DOI, विषय शब्द और Access mode सहित है, तो उपयोगकर्ता उसे आसानी से खोज सकता है।

अतः कहा जा सकता है कि **उच्च गुणवत्ता वाला ग्रंथसूची विवरण, सफल सूचना पुनर्प्राप्ति की कुंजी है।**

6.17 लघु प्रश्न एवं MCQ

A) लघु प्रश्न

1. Bibliographic Control क्या है?
2. UBC का मुख्य उद्देश्य क्या है?
3. ISBD द्वारा prescribed punctuation का क्या महत्व है?
4. Adequacy और Relevance में अंतर लिखिए।
5. OPAC में bibliographic description का योगदान लिखिए।
6. डिजिटल संसाधनों में description कठिन क्यों है?

B) MCQ— उत्तर सहित

1. ISBD का मुख्य उद्देश्य है—
   (A) वर्गीकरण (B) **compatible descriptive cataloguing worldwide** (C) circulation (D) shelving
   ✅उत्तर: (B)
2. ISBD का उद्देश्य records को किस रूप में मदद करना है?
   (A) केवल print (B) **electronic form conversion** (C) binding (D) weeding
   ✅उत्तर: (B)
3. Universal Bibliographic Control का संबंध मुख्यतः किससे है?
   (A) shelving (B) **records sharing & national responsibility** (C) stock verification (D) budgeting
   ✅उत्तर: (B)

4. Principle of Uniformity का उद्देश्य है—
   (A) अलग-अलग रिकॉर्ड बनाना (B) **एकरूप रिकॉर्ड बनाना** (C) केवल titles लिखना (D) random entry
   ✅उत्तर: (B)
5. Bibliographic description में prescribed order का लाभ है—
   (A) भ्रम बढ़ता है (B) **interchangeability बढ़ती है** (C) cost बढ़ती है (D) retrieval घटता है
   ✅उत्तर: (B)
6. ISBD mainly किसके लिए standard है?
   (A) archives only (B) museums only (C) **library community and beyond** (D) bookstores only
   ✅उत्तर: (C)
7. Identification principle मुख्यतः किस प्रश्न का उत्तर देता है?
   (A) कहाँ रखें (B) **यह दस्तावेज़ कौन-सा है?** (C) कितना खरीदा (D) किसने issue किया
   ✅उत्तर: (B)
8. OPAC में खोज के लिए सबसे महत्वपूर्ण है—
   (A) cover design (B) **structured bibliographic metadata** (C) shelf number only (D) accession only
   ✅उत्तर: (B)
9. ISBD में elements को अलग करने हेतु प्रयोग होता है—
   (A) emojis (B) **punctuation as delimiters** (C) colors (D) stamps
   ✅उत्तर: (B)
10. Adequacy का अर्थ है—
    (A) बहुत अधिक सूचना (B) **जितनी आवश्यक उतनी सूचना** (C) कोई सूचना नहीं (D) random note
    ✅उत्तर: (B)

6.18 अभ्यास प्रश्नों के उत्तर

Q. ISBD records को interchangeable कैसे बनाता है?
उत्तर (संक्षेप मॉडल): ISBD bibliographic description के तत्व निर्धारित करता है, उन्हें एक fixed order में प्रस्तुत करता है और punctuation द्वारा अलग करता है, जिससे विभिन्न देशों/स्रोतों के records एक-दूसरे के catalogues में आसानी से स्वीकार किए जा सकें।

Q. Bibliographic description और retrieval का संबंध लिखिए।
उत्तर: जब रिकॉर्ड में author/title/subject/standard numbers जैसे access points सटीक होते हैं, तो OPAC/डेटाबेस retrieval में precision और recall बेहतर होते हैं; गलत/अधूरे रिकॉर्ड retrieval failure का कारण बनते हैं।

6.19 संदर्भ ग्रंथ सूची

1. **IFLA.** International Standard Bibliographic Description (ISBD). Berlin: De Gruyter Saur.
2. **IFLA.** Statement of International Cataloguing Principles (ICP). *Anglo-American* Cataloguing Rules, 2nd ed. (AACR-2). Chicago: ALA.
3. **RDA Toolkit.** Resource Description and Access. American Library Association.
4. **Tillett, Barbara B.** What Is FRBR? A Conceptual Model for the Bibliographic Universe. Library of Congress.
5. **Bowers, Fredson.** Principles of Bibliographical Description. Princeton University Press.
6. **Gorman, Michael & Winkler, Paul.** Anglo-American Cataloguing Rules Explained. ALA.
7. **Taylor, Arlene G. & Joudrey, Daniel N.** The Organization of Information. Libraries Unlimited.
8. **Hider, Philip.** Information Resource Description: Creating and Managing Metadata. Facet Publishing.
9. **Chan, Lois Mai & Salaba, Athena.** Cataloging and Classification: An Introduction. Rowman & Littlefield.
10. **Kumar, P. S. G.** Theory and Practice of Cataloguing. B. R. Publishing.
11. **Lancaster, F. W.** Information Retrieval Systems. Wiley.
12. **रंगनाथन, एस. आर.** पुस्तकालय वर्गीकरण एवं सूचीकरण. मद्रास: सारदा रंगनाथन एंडोमेंट.
13. **शर्मा, जे. एन.** पुस्तकालय संगठन एवं प्रबंधन. नई दिल्ली: अटलांटिक पब्लिशर्स.
14. **कुमार, पी. एस. जी.** सूचना संगठन एवं सूचीकरण. नई दिल्ली: बी. आर. पब्लिशिंग.
15. **मिश्र, बी. एन.** ग्रंथसूची संगठन के सिद्धांत. वाराणसी: विश्वविद्यालय प्रकाशन.
16. **सिंह, एस. पी.** सूचना संगठन एवं पुनर्प्राप्ति. नई दिल्ली: एस्सार पब्लिकेशन.
17. **शुक्ला, आर. के.** पुस्तकालय सूचीकरण के सिद्धांत एवं व्यवहार. प्रयागराज: छात्र साहित्य.
18. **गुप्ता, डी. के.** आधुनिक पुस्तकालय एवं सूचना सेवाएँ. जयपुर: राज पब्लिशिंग.
19. **IGNOU.** *MLIS Study Material: Cataloguing and Bibliographic Control.* नई दिल्ली.
20. **उत्तराखंड / उत्तर प्रदेश / अन्य ODL विश्वविद्यालय.** *MLIS अध्ययन सामग्री: ग्रंथसूची विवरण एवं मानकीकरण.*

21. **राष्ट्रीय शैक्षिक अनुसंधान एवं प्रशिक्षण परिषद (NCERT) / UGC समर्थित सामग्री.** सूचना संगठन एवं ग्रंथसूची नियंत्रण.
22. **विश्वविद्यालय अनुदान आयोग (UGC).** ODL एवं उच्च शिक्षा में पुस्तकालय सेवाएँ.
23. **भारतीय पुस्तकालय संघ (ILA).** भारतीय पुस्तकालय प्रणाली एवं मानकीकरण पर लेख.

इकाई–7 : ग्रंथसूची विवरण के नियम (Rules for Bibliographic Description)

इकाई की रूपरेखा

7.1 प्रस्तावना

7.2 उद्देश्य

7.3 ग्रंथसूची नियमों की आवश्यकता

7.4 सूचीकरण नियमों की सामान्य विशेषताएँ

7.5 प्रमुख ग्रंथसूची नियमों का परिचय

  7.5.1 Panizzi's Rules

  7.5.2 Cutter's Rules

  7.5.3 Anglo-American Cataloguing Rules (AACR & AACR-2)

7.6 AACR-2 : संरचना एवं भाग

7.7 ग्रंथसूची विवरण के क्षेत्र (Areas of Description)

  7.7.1 शीर्षक एवं उत्तरदायित्व क्षेत्र

  7.7.2 संस्करण क्षेत्र

  7.7.3 प्रकाशन, वितरण क्षेत्र

  7.7.4 भौतिक विवरण क्षेत्र

  7.7.5 श्रृंखला क्षेत्र

7.8 विराम चिह्नों (Punctuation) का महत्व

7.9 विभिन्न प्रकार की सामग्री का ग्रंथसूची विवरण

7.9.1 पुस्तकें

7.9.2 पत्रिकाएँ

7.9.3 गैर-पुस्तक सामग्री

7.10 नियमों की सीमाएँ एवं आलोचना

7.11 RDA की ओर संक्रमण का संकेत

7.12 व्यावहारिक उदाहरण: AACR-2 के अनुसार विवरण

7.13 सारांश

7.14 शब्दावली

7.15 निबंधात्मक प्रश्न

7.16 लघु प्रश्न एवं MCQ

7.17 अभ्यास प्रश्नों के उत्तर

7.18 संदर्भ ग्रंथ सूची

7.1 प्रस्तावना (Introduction)

पुस्तकालयों में उपलब्ध सूचना संसाधनों को सुव्यवस्थित एवं उपयोगी बनाने के लिए **ग्रंथसूची विवरण (Bibliographic Description)** का विशेष महत्व है। ग्रंथसूची विवरण वह प्रक्रिया है, जिसके माध्यम से किसी दस्तावेज़ के महत्वपूर्ण पहचान तत्वों—जैसे शीर्षक, लेखक, संस्करण, प्रकाशन विवरण आदि—को निश्चित नियमों के अनुसार दर्ज किया जाता है।

जैसे-जैसे पुस्तकालय संग्रह का आकार बढ़ता है और विभिन्न प्रकार की सामग्री (पुस्तकें, पत्रिकाएँ, ई-संसाधन, AV सामग्री) जुड़ती जाती है, वैसे-वैसे **एकरूप और मानकीकृत विवरण** की आवश्यकता और भी बढ़ जाती है। इसी आवश्यकता की पूर्ति के लिए विभिन्न **ग्रंथसूची नियम (Bibliographic Rules)** विकसित किए गए हैं।

संक्षेप में, ग्रंथसूची विवरण के नियम पुस्तकालय उपयोगकर्ताओं को सूचना की **पहचान, खोज और चयन** में सहायता प्रदान करते हैं।

7.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के बाद विद्यार्थी—

1. ग्रंथसूची विवरण एवं नियमों की अवधारणा को समझ सकेंगे
2. ग्रंथसूची नियमों की आवश्यकता एवं महत्व स्पष्ट कर सकेंगे
3. सूचीकरण नियमों की सामान्य विशेषताओं को पहचान सकेंगे
4. प्रमुख ग्रंथसूची नियमों—Panizzi, Cutter एवं AACR-2—का तुलनात्मक अध्ययन कर सकेंगे
5. AACR-2 की संरचना एवं विवरण क्षेत्रों को समझ सकेंगे
6. विभिन्न प्रकार की सामग्री के लिए उपयुक्त ग्रंथसूची विवरण तैयार कर सकेंगे
7. ग्रंथसूची नियमों की सीमाओं एवं भविष्य की दिशा (RDA की ओर संक्रमण) को समझ सकेंगे

7.3 ग्रंथसूची नियमों की आवश्यकता (Need for Bibliographic Rules)

ग्रंथसूची नियमों की आवश्यकता इसलिए उत्पन्न हुई क्योंकि बिना मानकीकरण के तैयार किया गया विवरण **अस्पष्ट, असंगत और भ्रमपूर्ण** हो सकता है। विभिन्न पुस्तकालयों में यदि एक ही पुस्तक का विवरण अलग-अलग ढंग से किया जाए, तो सूचना का आदान-प्रदान एवं उपयोग कठिन हो जाता है।

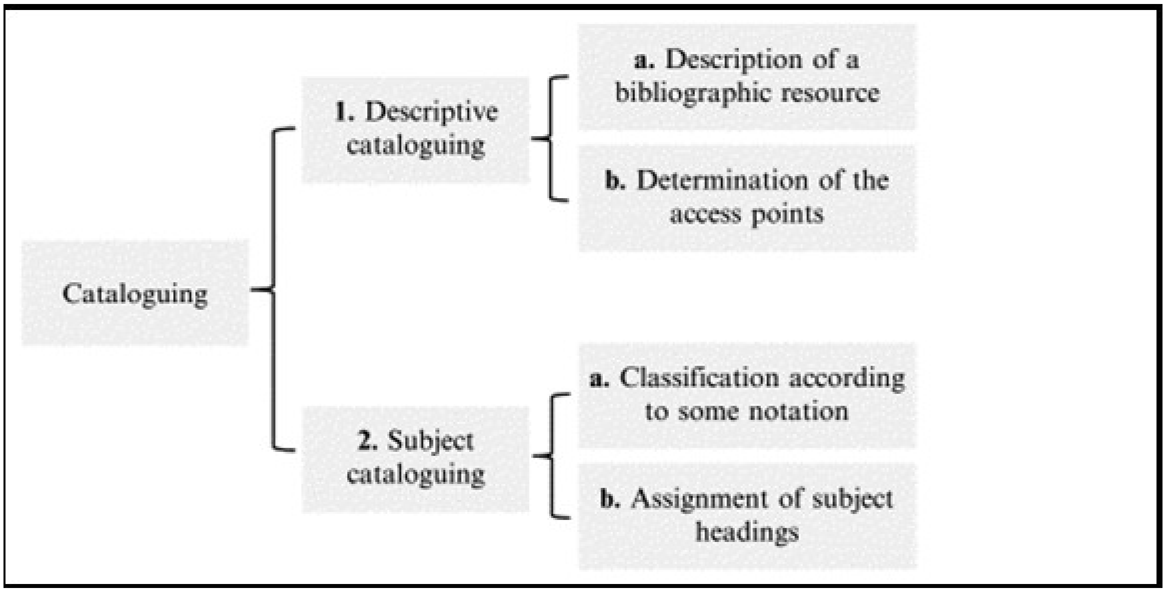
ग्रंथसूची नियमों की आवश्यकता के प्रमुख कारण निम्नलिखित हैं—

- **एकरूपता (Uniformity):** सभी दस्तावेज़ों का विवरण एक समान ढाँचे में प्रस्तुत करना
- **सटीक पहचान:** समान शीर्षक या लेखक वाली पुस्तकों में भेद स्पष्ट करना
- **सूचना पुनर्प्राप्ति में सहायता:** उपयोगकर्ताओं को वांछित सामग्री शीघ्र उपलब्ध कराना
- **डेटा साझा करने में सुविधा:** विभिन्न पुस्तकालयों एवं नेटवर्क के बीच सहयोग
- **डिजिटल वातावरण में अनुकूलन:** OPAC, डिजिटल लाइब्रेरी एवं रिपॉजिटरी के लिए मानकीकृत रिकॉर्ड

उदाहरण:
यदि किसी पुस्तक का प्रकाशन वर्ष या संस्करण सही ढंग से दर्ज न किया जाए, तो शोधार्थी गलत संस्करण का उपयोग कर सकता है। ग्रंथसूची नियम इस प्रकार की त्रुटियों को रोकते हैं।

7.4 सूचीकरण नियमों की सामान्य विशेषताएँ (General Features of Cataloguing Rules)

सूचीकरण नियम वे मानक सिद्धांत एवं दिशानिर्देश हैं, जिनके अनुसार किसी दस्तावेज़ का **ग्रंथसूची विवरण** तैयार किया जाता है। विभिन्न समयों में विकसित नियमों में कुछ **सामान्य एवं मूलभूत विशेषताएँ** समान रूप से पाई जाती हैं, जो उन्हें व्यावहारिक और उपयोगी बनाती हैं।



Diagram–1 : Cataloguing के प्रमुख घटक

व्याख्या: यह चित्र सूचीकरण (Cataloguing) के दो मुख्य भागों—**Descriptive Cataloguing** तथा **Subject Cataloguing**—को दर्शाता है। Descriptive Cataloguing में दस्तावेज़ का ग्रंथसूची विवरण एवं access points निर्धारित किए जाते हैं, जबकि Subject Cataloguing में वर्गीकरण एवं विषय शीर्षक प्रदान किए जाते हैं।

सूचीकरण नियमों की प्रमुख सामान्य विशेषताएँ निम्नलिखित हैं—

1. **एकरूपता (Uniformity):** सभी दस्तावेज़ों का विवरण एक समान ढाँचे और क्रम में प्रस्तुत किया जाता है, जिससे रिकॉर्ड समझने और साझा करने में सुविधा होती है।
2. **स्पष्टता एवं सटीकता (Clarity & Accuracy):** नियमों का उद्देश्य दस्तावेज़ की पहचान को स्पष्ट बनाना है, ताकि समान शीर्षक या लेखक वाली कृतियों में भ्रम न हो।

3. **उपयोगकर्ता-केंद्रित दृष्टिकोण:** सूचीकरण नियम इस प्रकार बनाए जाते हैं कि उपयोगकर्ता आसानी से वांछित सामग्री खोज और पहचान सके।
4. **मानकीकरण (Standardisation):** निर्धारित विराम-चिह्नों, क्रम और शब्दावली का प्रयोग कर विवरण को मानकीकृत किया जाता है।
5. **लचीलापन (Flexibility):** नियमों में नई सामग्री और माध्यमों (जैसे ई-संसाधन, AV सामग्री) के अनुसार विस्तार की क्षमता होती है।
6. **अंतरराष्ट्रीय उपयोगिता:** आधुनिक सूचीकरण नियम विभिन्न देशों एवं भाषाओं में उपयोग के लिए उपयुक्त होते हैं।

संक्षेप में, सूचीकरण नियम पुस्तकालय रिकॉर्ड को **विश्वसनीय, सुसंगत और उपयोगी** बनाते हैं।

7.5 प्रमुख ग्रंथसूची नियमों का परिचय (Introduction to Major Bibliographic Rules)

ग्रंथसूची विवरण के क्षेत्र में समय-समय पर अनेक नियम विकसित किए गए, जिनमें से कुछ नियमों ने आधुनिक सूचीकरण की नींव रखी। प्रमुख नियमों का संक्षिप्त परिचय निम्नलिखित है—

7.5.1 Panizzi’s Rules

Panizzi’s Rules का निर्माण **Anthony Panizzi** द्वारा 1841 में ब्रिटिश म्यूज़ियम लाइब्रेरी के लिए किया गया। ये नियम आधुनिक सूचीकरण के प्रारंभिक और आधारभूत नियम माने जाते हैं।

मुख्य विशेषताएँ—

- लेखक-आधारित (Author-based) सूचीकरण पर बल
- एक ही लेखक की सभी कृतियों को एक साथ लाने की अवधारणा
- पुस्तक को उसकी **बौद्धिक कृति (work)** के रूप में पहचान

महत्त्व:
Panizzi’s Rules ने सूचीकरण को केवल सूची बनाने से आगे बढ़ाकर **बौद्धिक संगठन** का रूप दिया।

7.5.2 Cutter’s Rules

Cutter's Rules का विकास **Charles Ammi Cutter** ने 1876 में किया। इन नियमों में सूचीकरण के **उद्देश्यों को स्पष्ट रूप से परिभाषित** किया गया।

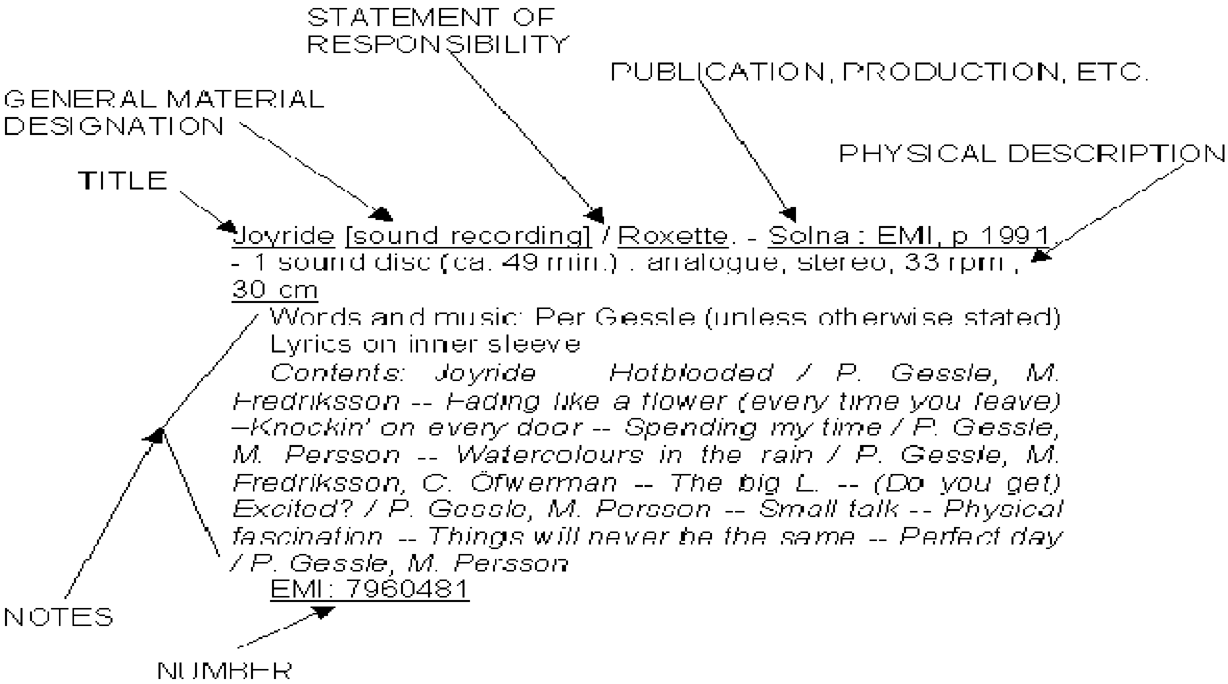
Cutter के प्रमुख उद्देश्य—

1. पुस्तक को लेखक, शीर्षक या विषय द्वारा खोजने में सहायता
2. उपयोगकर्ता को पुस्तक का चयन करने में मदद
3. पुस्तकालय संग्रह को सुव्यवस्थित ढंग से प्रदर्शित करना

महत्त्व:
Cutter's Rules ने सूचीकरण को स्पष्ट रूप से **user-oriented** बनाया।

7.5.3 Anglo-American Cataloguing Rules (AACR एवं AACR-2)

Anglo-American Cataloguing Rules (AACR) का प्रथम संस्करण **1967** में प्रकाशित हुआ, जबकि इसका संशोधित एवं विस्तारित रूप **AACR-2** वर्ष **1978** में विकसित किया गया। AACR-2 एक **internationally accepted cataloguing code** है, जिसका व्यापक उपयोग पुस्तकालयों एवं सूचना केंद्रों में किया जाता है।



मुख्य विशेषताएँ (Key Features)

- **Different types of materials** (पुस्तकें, पत्रिकाएँ, गैर-पुस्तक सामग्री) के लिए स्पष्ट नियम
- **ISBD-based areas of description**, जिससे विवरण एकरूप और मानकीकृत बनता है
- **OPAC एवं digital library environment** के लिए उपयुक्त संरचना
- **Access points** (main entry एवं added entries) पर विशेष ध्यान

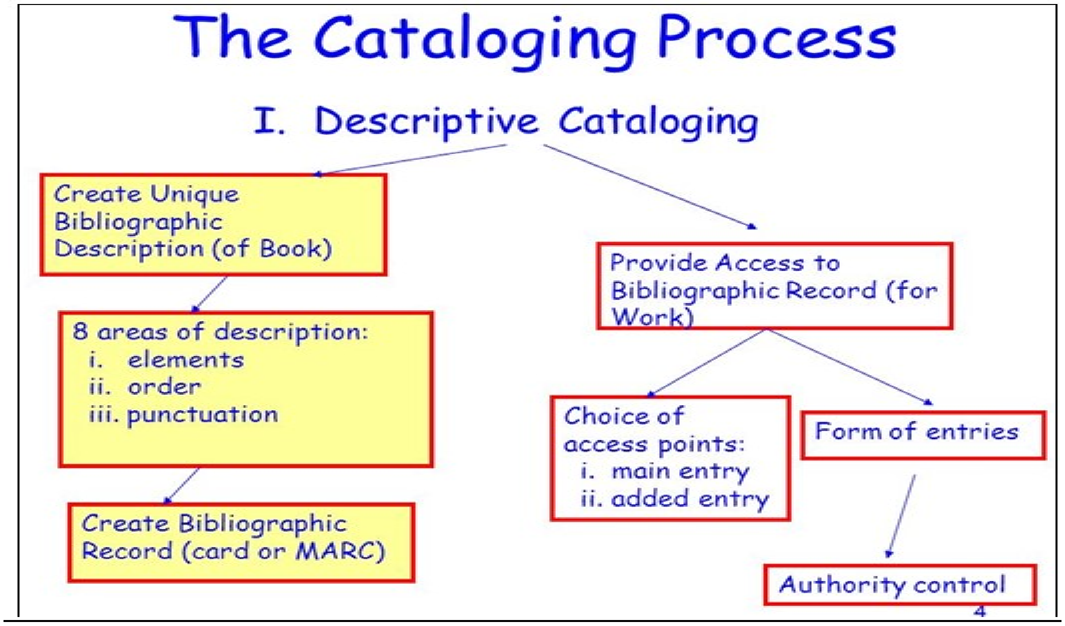
महत्त्व (Significance)

AACR-2 ने ग्रंथसूची विवरण को **practical, standardised और globally interoperable** बनाया तथा traditional card catalogue से **online catalogues (OPACs)** की ओर संक्रमण को सरल किया।

7.6 AACR-2 : संरचना एवं भाग (Structure and Parts of AACR-2)

AACR-2 (Anglo-American Cataloguing Rules, Second Edition) को दो मुख्य भागों में विभाजित किया गया है। यह विभाजन ग्रंथसूची विवरण को **systematic, standardised और practical** बनाता है।

Diagram–2 : Descriptive Cataloguing की प्रक्रिया



व्याख्या:
यह चित्र Descriptive Cataloguing की प्रक्रिया को स्पष्ट करता है, जिसमें **unique bibliographic description, 8 क्षेत्रों के अनुसार विवरण, access points का चयन** तथा **authority control** के माध्यम से पूर्ण bibliographic record (card या MARC) तैयार किया जाता है।

Part I : Description

यह भाग **Descriptive Cataloguing** से संबंधित है, जिसमें दस्तावेज़ का भौतिक एवं प्रकाशन संबंधी विवरण निश्चित नियमों के अनुसार दर्ज किया जाता है। इसमें निम्न बिंदुओं पर बल दिया जाता है—

- **Elements of description** (शीर्षक, लेखक, संस्करण, प्रकाशन आदि)
- **Order of elements** (तत्वों का निर्धारित क्रम)
- **Prescribed punctuation** (निर्धारित विराम-चिह्न)

Part II : Headings, Uniform Titles and References

यह भाग **Access points** से संबंधित है, जिससे उपयोगकर्ता दस्तावेज़ तक आसानी से पहुँच सके।
इसमें शामिल हैं—

- **Choice of headings** (main entry, added entries)
- **Form of headings**
- **Uniform titles**
- **Authority control**

महत्त्व:
AACR-2 की यह संरचना bibliographic records को **card catalogue से OPAC और MARC environment** तक सहज रूप से अनुकूल बनाती है।

7.7 ग्रंथसूची विवरण के क्षेत्र (Areas of Description) (Areas of Bibliographic Description as per AACR-2 / ISBD)

AACR-2 में ग्रंथसूची विवरण **ISBD (International Standard Bibliographic Description)** के अनुरूप विभिन्न क्षेत्रों (Areas) में किया जाता है। प्रत्येक क्षेत्र का **निर्धारित क्रम एवं punctuation** होता है।

ISBD Elements

```
│
├─ Title & Statement of Responsibility
│   └─ Title proper, Parallel title, Author/Editor
│
├─ Edition
│   └─ Version / Edition statement
│
├─ Material-Specific Area
│   └─ Map scale / Periodical numbering
│
└─ Publication / Distribution
    └─ Place, Publisher, Year
```

7.7.1 शीर्षक एवं उत्तरदायित्व क्षेत्र (Title and Statement of Responsibility Area)

इस क्षेत्र में दस्तावेज़ की **मुख्य पहचान (primary identification)** से संबंधित जानकारी दी जाती है। यह क्षेत्र उपयोगकर्ता को किसी कृति को **पहचानने (identify)** और **दूसरी कृतियों से भेद करने (distinguish)** में सहायता करता है।

इसमें निम्न तत्व सम्मिलित होते हैं—

- **Title proper** – दस्तावेज़ का मुख्य शीर्षक
- **Parallel title** – अन्य भाषा में दिया गया शीर्षक
- **Statement of responsibility** – लेखक, संपादक, अनुवादक आदि के नाम

यह क्षेत्र AACR-2 तथा ISBD के अंतर्गत ग्रंथसूची विवरण का **प्रथम और अत्यंत महत्वपूर्ण क्षेत्र** माना जाता है।

उदाहरण (Example):

Information Organization / **G. G. Chowdhury**

7.7.2 संस्करण क्षेत्र (Edition Area):

इस क्षेत्र में दस्तावेज़ के **संस्करण (edition / version)** का उल्लेख किया जाता है। यह क्षेत्र विशेष रूप से तब महत्त्वपूर्ण होता है जब किसी कृति के **एक से अधिक संस्करण** उपलब्ध हों, क्योंकि प्रत्येक संस्करण की सामग्री में अंतर हो सकता है।

उदाहरण (Example):
— 2nd ed.

7.7.3 प्रकाशन, वितरण क्षेत्र (Publication, Distribution Area)

इस क्षेत्र में दस्तावेज़ के प्रकाशन से संबंधित **bibliographic details** दिए जाते हैं, जिनसे उसके **प्रकाशक एवं प्रकाशन वर्ष** की पहचान होती है।

इसमें सम्मिलित होते हैं—

- **Place of publication**
- **Name of publisher**
- **Year of publication**

उदाहरण:
— New Delhi : Oxford University Press, 2019.

7.7.4 भौतिक विवरण क्षेत्र (Physical Description Area)

यह क्षेत्र दस्तावेज़ के **physical characteristics** को दर्शाता है, जिससे उपयोगकर्ता उसके आकार एवं स्वरूप का अनुमान लगा सके।

इसमें शामिल हैं—

- **Extent** (pages / volumes)
- **Illustrations**
- **Size**

उदाहरण:
— xii, 350 p. : ill. ; 24 cm.

7.7.5 श्रृंखला क्षेत्र (Series Area)

यदि कोई दस्तावेज़ किसी **series** का भाग हो, तो उस श्रृंखला का विवरण इस क्षेत्र में दिया जाता है। यह क्षेत्र संबंधित कृतियों को **आपस में जोड़ने (linking)** में सहायक होता है।

उदाहरण: — *(Library and Information Science Series ; 12)*

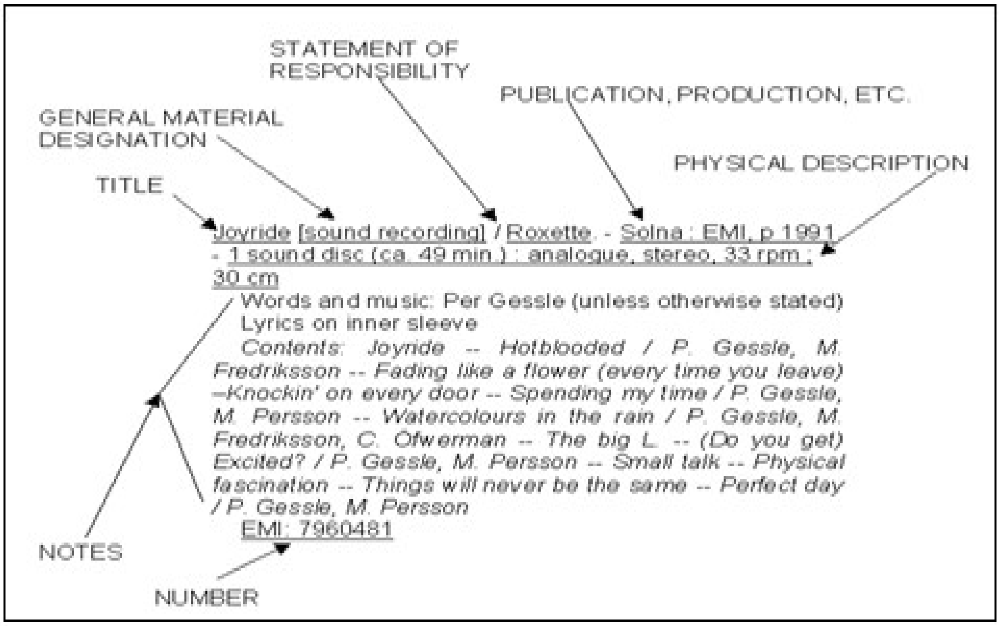
7.8 विराम-चिह्नों (Punctuation) का महत्व (Importance of Punctuation in AACR-2 / ISBD)

AACR-2 तथा ISBD में विराम-चिह्नों (punctuation) का प्रयोग केवल भाषा-सौंदर्य के लिए नहीं, बल्कि **bibliographic elements को स्पष्ट रूप से अलग-अलग दिखाने** के लिए किया जाता है। निर्धारित punctuation से ग्रंथसूची विवरण **uniform, machine-readable** और **internationally understandable** बनता है।

मुख्य बिंदु:

- विराम-चिह्न प्रत्येक **area of description** को अलग करता है
- OPAC एवं MARC records में **data parsing** को सरल बनाता है
- मानकीकरण (standardisation) सुनिश्चित करता है

उदाहरण: ***Information Organization* / G. G. Chowdhury. — 2nd ed. — New Delhi : OUP, 2019.**



चित्र 7.8 : ISBD में प्रयुक्त विराम-चिह्न एवं विवरण क्षेत्रों का क्रम

	3.1	Place of original production
		First place
;		*Subsequent place(s)
:	3.2	Name of producer
[]	3.3	Statement of function of producer
,	*3.4	Date of original production
--	3.5	Place of original distributor
:		Name of distributor
[]		Statement of function of distributor
,		Date of original distribution
--		*Place of additional company involved in production, distribution, etc.
:		Name of additional company involved in production, distribution, etc.
[]		Statement of function of additional company
,		*Date of additional company involvement

7.9 विभिन्न प्रकार की सामग्री का ग्रंथसूची विवरण (Bibliographic Description of Different Types of Materials)

AACR-2 में सामग्री के प्रकार के अनुसार विवरण में **थोड़ा-बहुत variation** पाया जाता है, ताकि प्रत्येक सामग्री का स्वरूप सही ढंग से व्यक्त किया जा सके।

7.9.1 पुस्तकें (Books)

पुस्तकों के लिए सभी मुख्य ISBD क्षेत्रों का उपयोग किया जाता है।

उदाहरण:
***Library Management* / R. K. Sharma. — New Delhi : ABC Publishers, 2018. — 250 p. ; 23 cm.**

7.9.2 पत्रिकाएँ (Serials / Journals)

पत्रिकाओं में **frequency, numbering** और **continuity** विशेष रूप से महत्त्वपूर्ण होती है।

उदाहरण:
***Journal of Library Science.* — Vol. 10, no. 2 (2022). — New Delhi : LIS Association.**

7.9.3 गैर-पुस्तक सामग्री (Non-Book Materials)

गैर-पुस्तक सामग्री में **material-specific details** (format, duration, medium) जोड़े जाते हैं।

उदाहरण:
***Digital Libraries* [DVD]. — New Delhi : IGNOU, 2020.**

7.10 नियमों की सीमाएँ एवं आलोचना (Limitations and Criticism of AACR-2)

यद्यपि **AACR-2** लंबे समय तक एक प्रभावी और मानकीकृत cataloguing code रहा है, फिर भी बदलते **digital** और **web-based information environment** में इसकी कुछ सीमाएँ स्पष्ट हुई हैं।

मुख्य सीमाएँ (Key Limitations):

- **Print-centric approach** — मुख्यतः भौतिक (physical) संसाधनों पर केंद्रित
- Digital एवं online resources के लिए **limited flexibility**
- **Complex rules & exceptions**, जिन्हें सीखना कठिन
- FRBR / relational data modelling को पूर्ण रूप से support नहीं करता
- Linked data और semantic web के लिए **inadequate structure**

आलोचना(Criticism):
AACR-2 को आधुनिक metadata standards और **machine-to-machine**

interoperability की आवश्यकताओं के अनुरूप नहीं माना गया, जिससे नए नियमों की आवश्यकता महसूस हुई।

7.11 RDA की ओर संक्रमण का संकेत (Shift from AACR-2 to RDA)

AACR-2 की सीमाओं के परिणामस्वरूप **RDA (Resource Description and Access)** का विकास किया गया। RDA एक **content standard** है, जिसे विशेष रूप से **digital, web और linked data environment** के लिए डिज़ाइन किया गया है।

RDA की प्रमुख विशेषताएँ:

- **FRBR / LRM-based conceptual model**
- Digital एवं non-book resources के लिए बेहतर अनुकूलता
- **Element-based structure**, जिससे reuse और interoperability संभव
- **Semantic web एवं linked data** के लिए उपयुक्त

संकेत (Indication):
AACR-2 से RDA की ओर संक्रमण यह दर्शाता है कि cataloguing अब केवल record creation नहीं, बल्कि **data modelling और resource discovery** का कार्य बन चुका है।

7.12 व्यावहारिक उदाहरण: AACR-2 के अनुसार विवरण(Practical Example of AACR-2 Based Description)

नीचे एक पुस्तक का **AACR-2 एवं ISBD** के अनुसार संक्षिप्त ग्रंथसूची विवरण प्रस्तुत है—

उदाहरण:
***Information Organization* / G. G. Chowdhury. — 2nd ed. — London : Facet Publishing, 2004. — xii, 350 p. ; 24 cm.**

यह उदाहरण दर्शाता है कि AACR-2 में—

- निर्धारित **areas of description**
- **order of elements**
- एवं **prescribed punctuation**
  का पालन अनिवार्य है।

7.13 सारांश (Summary)

इस इकाई में ग्रंथसूची विवरण के नियमों (Rules for Bibliographic Description) का क्रमबद्ध और व्यावहारिक अध्ययन किया गया। इकाई की शुरुआत सूचीकरण नियमों की आवश्यकता एवं सामान्य विशेषताओं से हुई, जिसके अंतर्गत **Panizzi's Rules, Cutter's Rules तथा AACR एवं AACR-2** जैसे प्रमुख ग्रंथसूची नियमों का परिचय दिया गया।

AACR-2 की संरचना एवं उसके दो भागों—**Part I (Description)** तथा **Part II (Access Points and Authority Control)**—को स्पष्ट किया गया। इसके अंतर्गत **ISBD आधारित ग्रंथसूची विवरण के विभिन्न क्षेत्र**, जैसे शीर्षक एवं उत्तरदायित्व क्षेत्र, संस्करण क्षेत्र, प्रकाशन एवं वितरण क्षेत्र, भौतिक विवरण क्षेत्र तथा श्रृंखला क्षेत्र का उदाहरण सहित वर्णन किया गया।

इकाई में विराम-चिह्नों (punctuation) के महत्व, विभिन्न प्रकार की सामग्री (पुस्तक, पत्रिका एवं गैर-पुस्तक सामग्री) के ग्रंथसूची विवरण, AACR-2 की सीमाओं एवं आलोचनाओं तथा **RDA की ओर संक्रमण** पर भी प्रकाश डाला गया। अंत में AACR-2 के अनुसार एक व्यावहारिक उदाहरण प्रस्तुत किया गया।

संक्षेप में, यह इकाई यह स्पष्ट करती है कि मानकीकृत ग्रंथसूची विवरण पुस्तकालयों में **सटीक पहचान, प्रभावी सूचना-पुनर्प्राप्ति और अंतरराष्ट्रीय संसाधन-साझेदारी** के लिए अनिवार्य है।

7.14 शब्दावली (Glossary)

ग्रंथसूची विवरण (Bibliographic Description): दस्तावेज़ की पहचान हेतु उसके शीर्षक, लेखक, संस्करण, प्रकाशन, भौतिक स्वरूप आदि का मानकीकृत विवरण।

सूचीकरण (Cataloguing): पुस्तकालय संसाधनों का व्यवस्थित वर्णन एवं अभिलेखन करने की प्रक्रिया।

AACR (Anglo-American Cataloguing Rules): अंग्रेज़ी-भाषी देशों में विकसित एक प्रमुख सूचीकरण नियमावली।

AACR-2: AACR का संशोधित संस्करण, जो ISBD के अनुरूप मानकीकृत ग्रंथसूची विवरण प्रदान करता है।

ISBD (International Standard Bibliographic Description): IFLA द्वारा विकसित अंतरराष्ट्रीय मानक, जो ग्रंथसूची विवरण के क्षेत्रों, क्रम एवं विराम-चिह्नों को निर्धारित करता है।

Descriptive Cataloguing: दस्तावेज़ के भौतिक एवं प्रकाशन संबंधी विवरण को दर्ज करने की प्रक्रिया।

Access Points: वे प्रविष्टियाँ (जैसे लेखक, शीर्षक, विषय) जिनके माध्यम से उपयोगकर्ता दस्तावेज़ तक पहुँच प्राप्त करता है।

Statement of Responsibility: लेखक, संपादक, अनुवादक आदि के नामों का उल्लेख।

Edition Area: दस्तावेज़ के संस्करण से संबंधित विवरण का क्षेत्र।

Physical Description Area: दस्तावेज़ के पृष्ठों, आकार, चित्रों आदि का विवरण।

Series Area: किसी श्रृंखला का भाग होने पर श्रृंखला संबंधी जानकारी।

Punctuation: ग्रंथसूची विवरण में विभिन्न तत्वों को अलग-अलग दर्शाने हेतु प्रयुक्त निर्धारित विराम-चिह्न।

RDA (Resource Description and Access): AACR-2 के बाद विकसित आधुनिक content standard, जो digital एवं linked-data environment के लिए उपयुक्त है।

7.15 निबंधात्मक प्रश्न (Essay Type Questions):

प्रश्न 1.**ग्रंथसूची विवरण के नियमों की आवश्यकता एवं महत्व पर प्रकाश डालिए।**

उत्तर:
ग्रंथसूची विवरण के नियमों की आवश्यकता इसलिए होती है क्योंकि पुस्तकालयों में उपलब्ध संसाधनों की संख्या निरंतर बढ़ रही है। बिना मानकीकृत नियमों के दस्तावेज़ों का विवरण असंगठित हो जाएगा, जिससे सूचना-पुनर्प्राप्ति कठिन हो जाएगी। मानकीकृत नियम एकरूपता, स्पष्टता तथा अंतरराष्ट्रीय संसाधन-साझेदारी सुनिश्चित करते हैं। ये OPAC और डिजिटल लाइब्रेरी वातावरण में सटीक खोज (accurate retrieval) को संभव बनाते हैं।

प्रश्न 2. **Panizzi's Rules एवं Cutter's Rules का तुलनात्मक विवेचन कीजिए।**

उत्तर:
Panizzi's Rules ने लेखक-आधारित सूचीकरण और बौद्धिक कृति (work) की अवधारणा को स्थापित किया, जबकि Cutter's Rules ने सूचीकरण के उद्देश्यों को स्पष्ट कर इसे उपयोगकर्ता-केंद्रित बनाया।
Panizzi का दृष्टिकोण संग्रह-केंद्रित था, जबकि Cutter ने खोज, पहचान और चयन को प्राथमिकता दी। दोनों नियमों ने आधुनिक सूचीकरण के विकास में महत्वपूर्ण योगदान दिया।

प्रश्न 3. **AACR-2 की संरचना एवं उसके भागों का वर्णन कीजिए।**

उत्तर:
AACR-2 दो भागों में विभाजित है—Part I (Description) और Part II (Headings and Access Points)।
Part I में ISBD आधारित ग्रंथसूची विवरण के विभिन्न क्षेत्र निर्धारित किए गए हैं, जबकि Part II में headings, uniform titles और authority control से संबंधित नियम दिए गए हैं। यह संरचना OPAC और MARC environment के लिए उपयुक्त है।

प्रश्न 4. **ISBD आधारित ग्रंथसूची विवरण के क्षेत्रों का संक्षिप्त वर्णन कीजिए।**

उत्तर:
ISBD के अनुसार ग्रंथसूची विवरण विभिन्न क्षेत्रों में किया जाता है, जैसे शीर्षक एवं उत्तरदायित्व क्षेत्र, संस्करण क्षेत्र, प्रकाशन क्षेत्र, भौतिक विवरण क्षेत्र तथा श्रृंखला क्षेत्र। प्रत्येक क्षेत्र का निश्चित क्रम और punctuation होता है, जिससे अंतरराष्ट्रीय स्तर पर एकरूपता बनी रहती है।

प्रश्न 5. **AACR-2 की सीमाएँ स्पष्ट करते हुए RDA की आवश्यकता बताइए।**

उत्तर:
AACR-2 मुख्यतः print-centric है और डिजिटल संसाधनों के लिए सीमित लचीलापन प्रदान करता है।
इन सीमाओं के कारण RDA विकसित किया गया, जो FRBR आधारित, element-oriented तथा linked-data compatible है। RDA आधुनिक डिजिटल सूचना-परिवेश के लिए अधिक उपयुक्त है।

7.16 लघु प्रश्न (Short Answer Questions)

1. ग्रंथसूची विवरण से आप क्या समझते हैं?
2. Descriptive Cataloguing क्या है?

3. Edition Area का महत्व लिखिए।
4. ISBD में punctuation क्यों आवश्यक है?
5. Non-book materials से आप क्या समझते हैं?

7.16 MCQs (Multiple Choice Questions)

1. AACR-2 का प्रथम प्रकाशन वर्ष है—
   a) 1967
   b) 1971
   c) **1978 ✔**
   d) 1985

2. ISBD का विकास किस संस्था ने किया?
   a) UNESCO
   b) **IFLA ✔**
   c) ALA
   d) LC

3. Title proper किस क्षेत्र में आता है?
   a) Edition Area
   b) **Title & Statement of Responsibility Area ✔**
   c) Physical Description Area
   d) Series Area

4. Edition Area का उद्देश्य है—
   a) आकार बताना
   b) प्रकाशक बताना
   c) **संस्करण बताना ✔**
   d) विषय बताना

5. Physical Description Area में क्या दिया जाता है?
   a) लेखक
   b) प्रकाशन वर्ष
   c) **पृष्ठ, आकार ✔**
   d) विषय शीर्षक

6. Series Area का प्रयोग कब होता है?
   a) जब लेखक हो
   b) जब संस्करण हो
   c) **जब दस्तावेज़ श्रृंखला का भाग हो ✔**
   d) जब चित्र हों

7. AACR-2 का Part II संबंधित है—
   a) Description
   b) **Access points ✔**
   c) Binding
   d) Shelving

8. RDA आधारित है—
   a) MARC
   b) ISBD
   c) **FRBR ✔**
   d) OPAC

9. Journal किस श्रेणी में आता है?
   a) Book
   b) Non-book
   c) **Serial ✔**
   d) Map

10. Punctuation का मुख्य उद्देश्य है—
   a) सजावट
   b) भाषा सुधार
   c) **तत्वों को अलग दिखाना ✔**
   d) अनुवाद

7.17 अभ्यास प्रश्नों के उत्तर (Hints for Practice)

- AACR-2 = International cataloguing standard
- ISBD = Areas + order + punctuation
- RDA = Modern, digital-ready content standard
- OPAC records = AACR-2 / ISBD based

7.18 संदर्भ ग्रंथ सूची (References)

1. सिंह, के. पी. – सूचीकरण एवं वर्गीकरण
2. शर्मा, आर. – ग्रंथसूची संगठन
3. वर्मा, एस. – पुस्तकालय सूचीकरण
4. त्रिपाठी, बी. एन. – सूचना संगठन
5. IGNOU – Cataloguing and Classification SLM
6. यूजीसी-DEB अध्ययन सामग्री (LIS)
7. विश्वविद्यालय अनुदान आयोग – Library Science Modules

8. पाण्डेय, आर. – डिजिटल पुस्तकालय
9. मिश्रा, ए. – सूचना पुनर्प्राप्ति
10. राष्ट्रीय पुस्तकालय प्रशिक्षण दस्तावेज़
11. IFLA. ISBD Consolidated Edition
12. Taylor, A. G. Introduction to Cataloging and Classification
13. Chan, L. M. Cataloging and Classification
14. RDA Toolkit. Resource Description and Access
15. Library of Congress. Cataloging Policy Statements
16. Hider, P. Information Resource Description
17. Smiraglia, R. The Nature of "A Work"
18. Svenonius, E. The Intellectual Foundation of Information Organization
19. ALA. Cataloging Rules and Principles

इकाई–8 : ग्रंथसूची अभिलेखों के मानक (Standards for Bibliographic Records)

इकाई की रूपरेखा

8.1 प्रस्तावना

8.2 उद्देश्य

8.3 ग्रंथसूची मानकों की अवधारणा

8.4 मानकीकरण की आवश्यकता एवं लाभ

8.5 ISBD (International Standard Bibliographic Description)

8.5.1 उद्देश्य एवं विशेषताएँ

8.5.2 ISBD के क्षेत्र

8.6 FRBR : अवधारणा एवं घटक

8.6.1 Work, Expression, Manifestation, Item

8.7 FRAD एवं FRSAD का संक्षिप्त परिचय

8.8 RDA : पृष्ठभूमि एवं उद्देश्य

8.9 RDA बनाम AACR-2

8.10 अंतरराष्ट्रीय एवं राष्ट्रीय मानकों की भूमिका

8.11 डिजिटल पुस्तकालयों में मानकों का महत्व

8.12 व्यावहारिक उदाहरण: ISBD आधारित विवरण

8.13 वर्तमान प्रवृत्तियाँ एवं भविष्य

8.14 सारांश

8.15 शब्दावली

8.16 निबंधात्मक प्रश्न

8.17 लघु प्रश्न एवं MCQ

8.18 अभ्यास प्रश्नों के उत्तर

8.19 संदर्भ

8.20 संदर्भ ग्रंथ सूची

8.1 प्रस्तावना

आधुनिक पुस्तकालय अब केवल पुस्तकों का भौतिक संग्रह नहीं रह गए हैं, बल्कि वे **डिजिटल, स्वचालित तथा नेटवर्क-आधारित सूचना प्रणालियाँ** बन चुके हैं। वर्तमान परिदृश्य में एक ही सूचना संसाधन को **OPAC, Digital Library, Institutional Repository** तथा **Union Catalogue** जैसे विभिन्न प्लेटफॉर्मों पर एक साथ उपयोग किया जाता है।

ऐसे परिवेश में ग्रंथसूची अभिलेख (Bibliographic Records) तभी प्रभावी सिद्ध होते हैं जब वे **मानकीकृत (Standardized)** हों।
ग्रंथसूची मानक यह सुनिश्चित करते हैं कि—

✔ रिकॉर्ड **मानव द्वारा पढ़ने योग्य** हों
✔ रिकॉर्ड **मशीन द्वारा समझे और संसाधित** किए जा सकें
✔ विभिन्न पुस्तकालय प्रणालियों के बीच रिकॉर्ड **साझा (share)** किए जा सकें
✔ OPAC और डिजिटल डेटाबेस में **एकसमान एवं सटीक खोज** संभव हो सके

इस प्रकार, ग्रंथसूची मानक आधुनिक पुस्तकालयों में **सूचना संगठन, पुनर्प्राप्ति तथा संसाधन-साझाकरण** की आधारशिला के रूप में कार्य करते हैं।

8.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात शिक्षार्थी—

- ग्रंथसूची मानकों की अवधारणा समझ सकेंगे
- ISBD, MARC और Metadata standards की भूमिका जान सकेंगे
- Print और Digital resources के लिए मानकों का प्रयोग समझ सकेंगे

- आधुनिक पुस्तकालयों में मानकीकरण के महत्व को स्पष्ट कर सकेंगे

8.3 ग्रंथसूची मानकों की अवधारणा (Concept of Bibliographic Standards):

ग्रंथसूची मानक (Bibliographic Standards) वे नियम, ढाँचे एवं तकनीकी प्रारूप होते हैं,

ग्रंथसूची मानक वे स्वीकृत ढाँचे हैं जो यह निर्धारित करते हैं कि—

- ग्रंथसूची डेटा **कैसे संरचित** होगा
- **किस क्रम** में तत्व दिए जाएँगे
- रिकॉर्ड को **किस format** में संग्रहीत व साझा किया जाएगा

सरल शब्दों में

Rules = Content
Standards = Structure + Exchange

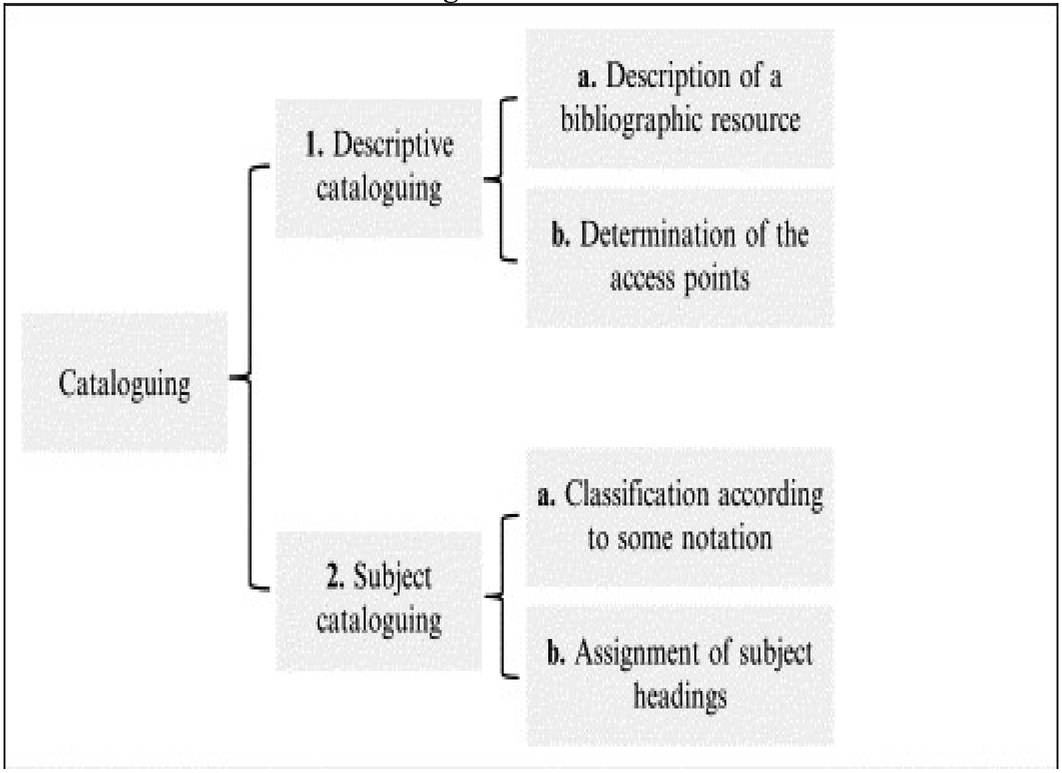


Figure: Components of Cataloguing: Descriptive and Subject Cataloguing

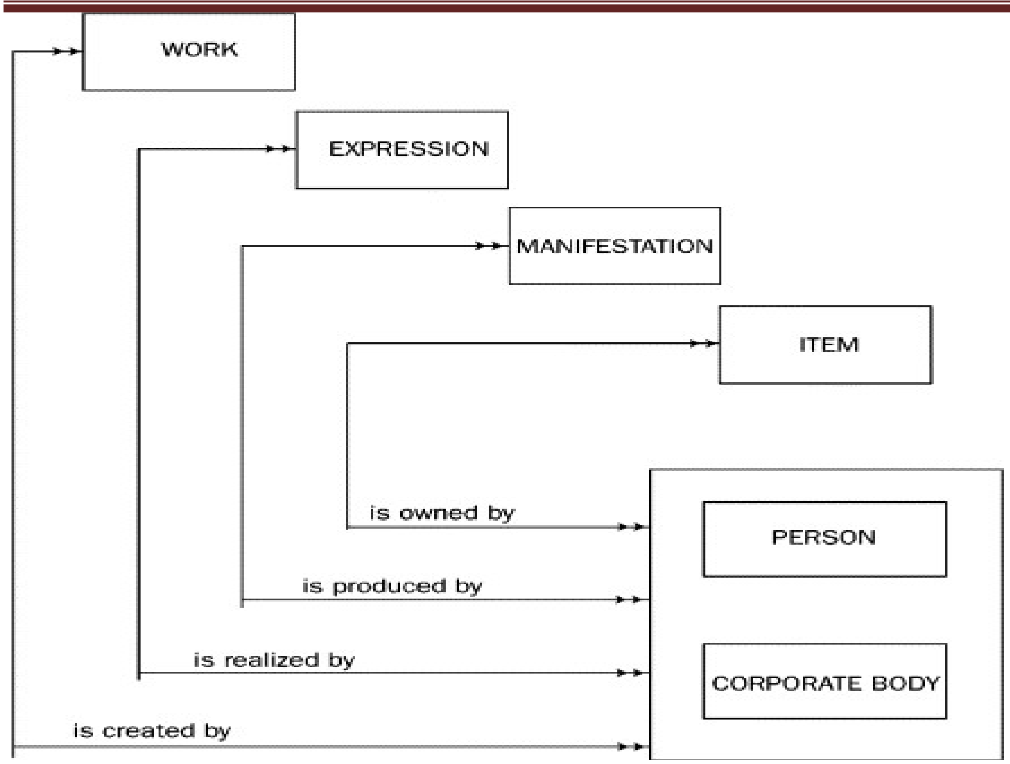


Figure: ***FRBR Entity–Relationship Model (Work–Expression–Manifestation–Item)***

व्याख्या: यह चित्र FRBR मॉडल को दर्शाता है, जिसमें किसी रचना के कार्य (Work), उसकी अभिव्यक्ति (Expression), प्रकाशन रूप (Manifestation) और उसकी प्रति (Item) के आपसी संबंध को समझाया गया है।

8.4 मानकीकरण की आवश्यकता एवं लाभ

आवश्यकता

- एकरूपता (Uniformity)
- Record Sharing
- Interoperability
- Automation support

लाभ

- Duplicate cataloguing में कमी
- समय और लागत की बचत

- तेज़ और सटीक सूचना पुनर्प्राप्ति

8.4 मानकीकरण की आवश्यकता एवं लाभ

मानकीकरण क्यों आवश्यक है?

- रिकॉर्ड की **एकरूपता (Uniformity)** के लिए
- **Record Sharing** एवं Resource Sharing हेतु
- राष्ट्रीय एवं अंतरराष्ट्रीय **सूचना सहयोग** के लिए
- Automation और OPAC search को प्रभावी बनाने हेतु

लाभ

- Duplicate cataloguing में कमी
- समय और लागत की बचत
- उपयोगकर्ता को सटीक और तेज़ खोज

8.5 ISBD (International Standard Bibliographic Description)

ISBD, **IFLA (International Federation of Library Associations and Institutions)** द्वारा विकसित एक **अंतरराष्ट्रीय वर्णनात्मक मानक (International Descriptive Standard)** है।
यह ग्रंथसूची विवरण (Bibliographic Description) के **तत्वों का क्रम (order of elements)** तथा **निर्धारित विराम-चिह्न (prescribed punctuation)** तय करता है, जिससे रिकॉर्ड **अंतरराष्ट्रीय स्तर पर समझने योग्य और साझा करने योग्य** बनते हैं।

ISBD के मुख्य उद्देश्य (Objectives of ISBD)

- अंतरराष्ट्रीय स्तर पर **एकसमान ग्रंथसूची विवरण (Uniform description)**
- **भाषा-बाधा (Language barrier)** को कम करना
- ग्रंथसूची रिकॉर्ड को **Machine-readable conversion** के लिए उपयुक्त बनाना
- रिकॉर्ड की **Interchangeability** को बढ़ाना

```
              3.1   Place of original production
                       First place
;                     *Subsequent place(s)
:             3.2   Name of producer
[]            3.3   Statement of function of
                      producer
,            *3.4   Date of original production
—             3.5   Place of original distributor
:                   Name of distributor
[]                  Statement of function of
                      distributor
,                   Date of original distribution
—                  *Place of additional company
                      involved in production,
                      distribution, etc.
:                   Name of additional company
                      involved in production,
                      distribution, etc.
[]                  Statement of function of
                      additional company
,                  *Date of additional company
                      involvement
```

<u>चित्र: ISBD के प्रकाशन एवं वितरण क्षेत्र (Publication and Distribution Area) के तत्व</u>

व्याख्या:

यह चित्र **ISBD** में प्रकाशन, उत्पादन और वितरण से संबंधित विभिन्न विवरणों जैसे स्थान, निर्माता और तिथि को क्रमबद्ध रूप में दर्शाता है।

ISBD के प्रमुख क्षेत्र (Main Areas of ISBD)

1. **Title & Statement of Responsibility**
2. **Edition Area**
3. **Material-specific Area**
4. **Publication / Distribution Area**
5. **Physical Description Area**
6. **Series Area**
7. **Notes Area**
8. **Standard Number Area (ISBN / ISSN)**

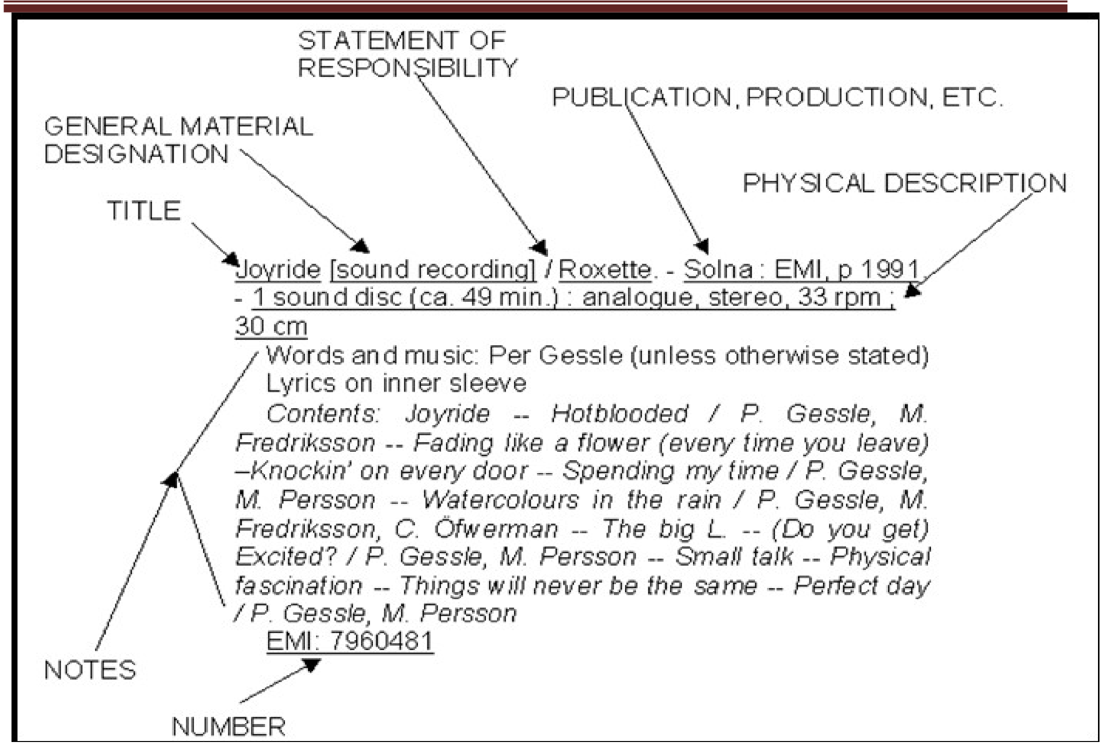


Fig.: Example of an ISBD Bibliographic Record)

व्याख्या :यह चित्र ISBD के प्रमुख क्षेत्रों तथा विराम-चिह्नों के प्रयोग को स्पष्ट करता है।

8.6 MARC (Machine Readable Cataloguing)

MARC एक **डेटा संचार मानक (Data Communication Standard)** है, जिसे Library of Congress द्वारा विकसित किया गया।

8.6 MARC (Machine Readable Cataloguing)

MARC एक **Data Communication Standard** है, जिसे **Library of Congress (USA)** द्वारा विकसित किया गया। यह ग्रंथसूची डेटा को इस प्रकार कोडित (encode) करता है कि **कंप्यूटर उसे पढ़, समझ और साझा कर सके।**

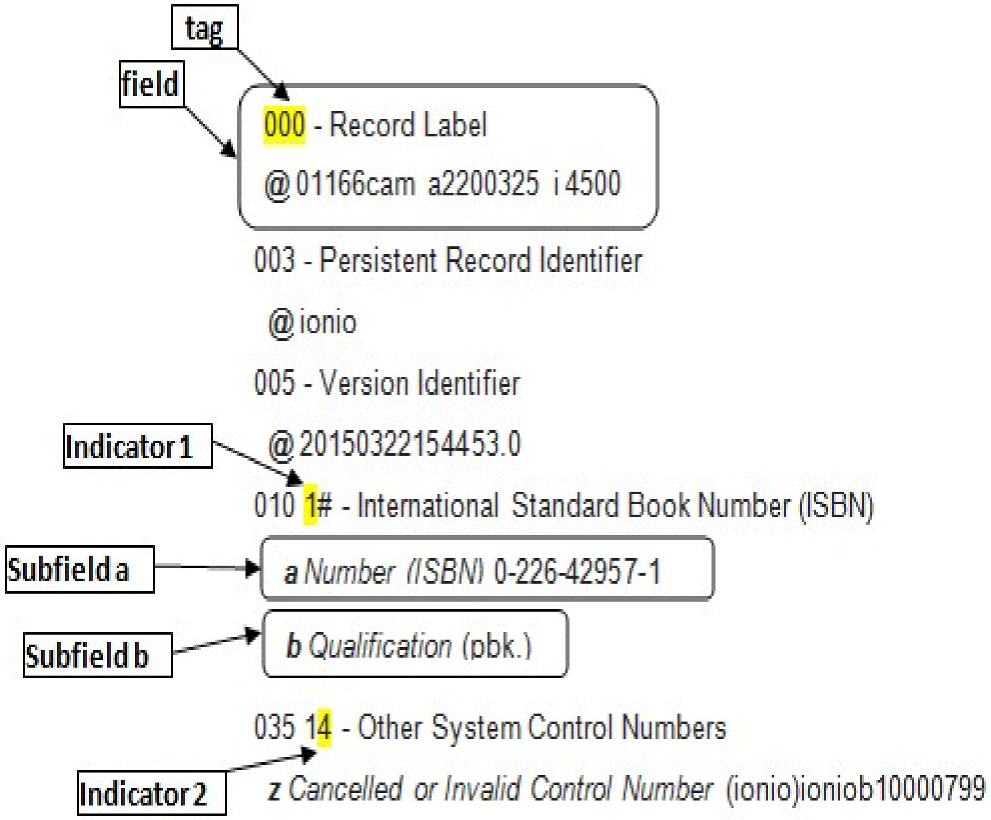
MARC का अर्थ (Machine Readable Catalogue)

MARC = ऐसा format जिसमें bibliographic data computer-readable होता है।

- ✓ MARC **cataloguing rules नहीं बताता**, बल्कि रिकॉर्ड को **machine-processable** बनाता है।

MARC का अर्थ

MARC = **ऐसा फॉर्मेट जिसमें कंप्यूटर ग्रंथसूची डेटा को पढ़, समझ और साझा** कर सके।



MARC रिकॉर्ड की संरचना (Structure of MARC Record)

- **Tag** – Field की पहचान (जैसे 100, 245, 260)
- **Indicators** – Field interpretation को नियंत्रित करते हैं
- **Subfields** – Data elements (‡a, ‡b, ‡c)

COMMONLY USED TAGS

TAG	DESIGNATION
010	Library of Congress Control Number
020	ISBN
100	Personal name main entry (Author)
245	Title and statement of responsibility
250	Edition statement
260	Imprint (Publisher, place and date)
300	Physical description
490	Series information
650	Topical subject heading
700	Personal name added entries (Co-authors, editors, illustrator etc.)

उदाहरण :

245 ‡a Knowledge Organization ‡b theory and practice ‡c S. R. Sharma

8.7 प्रमुख MARC प्रारूप (Major MARC Formats)

MARC (Machine Readable Cataloguing) के विभिन्न प्रारूप विकसित किए गए हैं, जिनका उद्देश्य ग्रंथसूची रिकॉर्ड को **machine-readable** एवं **exchange-friendly** बनाना है।

- **MARC 21** – सबसे अधिक प्रचलित एवं वर्तमान में अंतरराष्ट्रीय स्तर पर व्यापक रूप से उपयोग किया जाने वाला प्रारूप
- **UNIMARC** – **IFLA** द्वारा समर्थित, विभिन्न देशों के MARC प्रारूपों के बीच interoperability हेतु विकसित
- **INDMARC** – भारत में विकसित MARC प्रारूप (अब सीमित प्रयोग में)

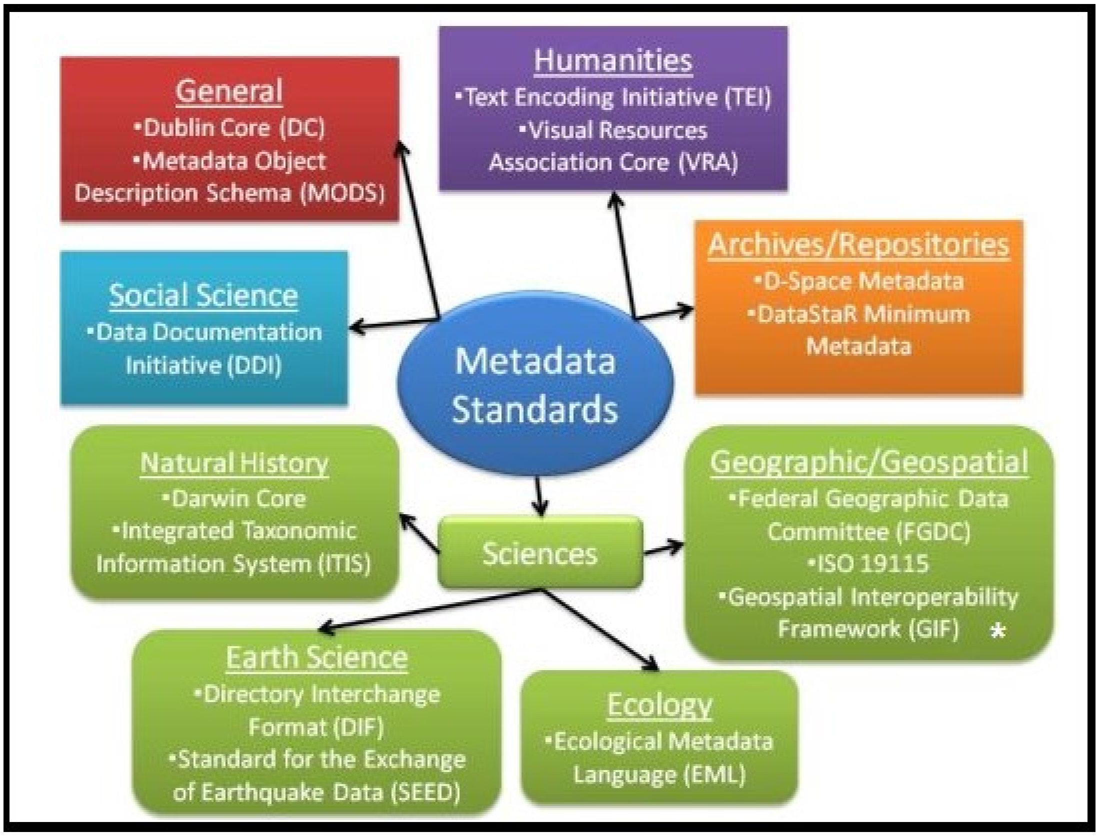
✓ **वर्तमान में अधिकांश आधुनिक पुस्तकालय MARC 21 का ही प्रयोग करते हैं।**

8.8 Metadata मानक (Metadata Standards)

Metadata का अर्थ है—
"Data about Data" अर्थात किसी संसाधन के बारे में संरचित जानकारी।

डिजिटल पुस्तकालयों (Digital Libraries), रिपॉज़िटरी (Repositories) और वेब संसाधनों (Web Resources) में Metadata का प्रयोग—

- संसाधनों की **पहचान (Identification)**
- **खोज (Discovery)**
- **प्रवेश (Access)**
- एवं **दीर्घकालिक संरक्षण (Preservation)**
  के लिए किया जाता है।

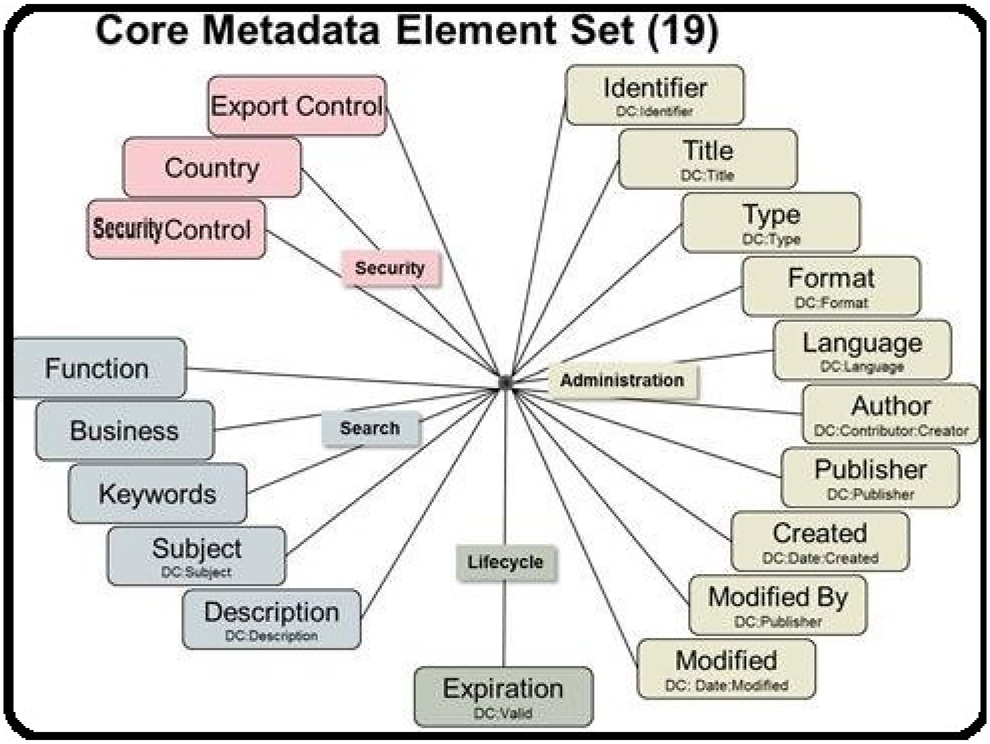


Dublin Core Metadata Standard

Dublin Core एक **सरल, लचीला एवं web-friendly metadata standard** है, जिसका प्रयोग डिजिटल संसाधनों के वर्णन हेतु व्यापक रूप से किया जाता है।

इसके **15 core elements** होते हैं, जिनमें प्रमुख हैं—

- **Title**
- **Creator**
- **Subject**
- **Date**
- **Format**
- **Identifier**



यह चित्र Dublin Core के प्रमुख मेटाडाटा तत्वों को दर्शाता है, जिनका प्रयोग डिजिटल संसाधनों की पहचान, खोज एवं वेब-आधारित सूचना पुनर्प्राप्ति को सरल बनाने के लिए किया जाता है। यह सरल, लचीला और web-friendly मानक है।

8.9 अन्य महत्वपूर्ण मानक (Other Important Standards)

डिजिटल एवं नेटवर्क-आधारित पुस्तकालय परिवेश में कुछ सहायक मानकों (supporting standards) की भी महत्वपूर्ण भूमिका होती है—

- **ISBN / ISSN (International Standard Book / Serial Number)** – पुस्तकों एवं धारावाहिक प्रकाशनों की **Unique Identification** हेतु प्रयुक्त मानक संख्या।
- **Z39.50** – विभिन्न डेटाबेस एवं पुस्तकालय प्रणालियों के बीच **Search & Retrieval Protocol**, जिससे एक सिस्टम से दूसरे सिस्टम में खोज संभव होती है।
- **OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)** – विभिन्न डिजिटल रिपॉज़िटरी से मेटाडाटा के **संग्रह (harvesting)** हेतु प्रयुक्त मानक।

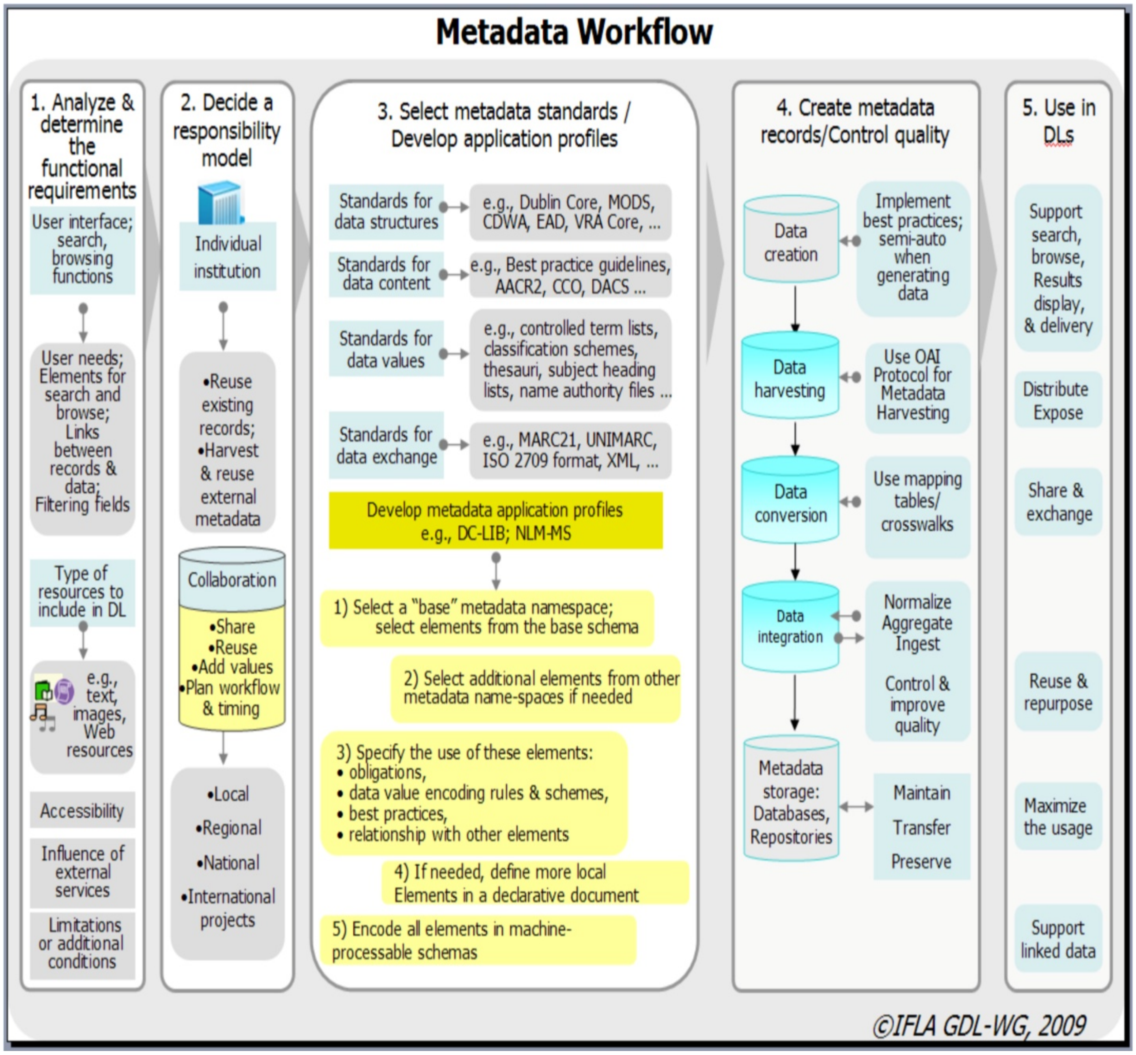
Imp. Points:

Z39.50 = Search & Retrieval

OAI-PMH = Metadata Harvesting

8.10 विभिन्न मानकों की तुलनात्मक भूमिका (Comparative Role of Standards)

मानक (Standard)	मुख्य उद्देश्य (Main Role)
ISBD	ग्रंथसूची विवरण का क्रम एवं निर्धारित punctuation
MARC	ग्रंथसूची डेटा को **machine-readable** बनाना
Metadata Standards	डिजिटल संसाधनों का संरचित वर्णन
ISBN / ISSN	संसाधनों की **Unique** पहचान



यह चित्र ISBD, MARC और Metadata मानकों की संयुक्त भूमिका को दर्शाता है, जिनके माध्यम से ग्रंथसूची विवरण को मानकीकृत, मशीन-पठनीय एवं डिजिटल परिवेश के अनुकूल बनाया जाता है।

8.11 डिजिटल पुस्तकालयों में मानकों का महत्व (Importance of Standards in Digital Libraries)

डिजिटल परिवेश में यदि मानकों (Standards) का प्रयोग न किया जाए, तो—

- **Interoperability** संभव नहीं हो पाती
- संसाधनों की **Global visibility** कम हो जाती है

- **Long-term preservation** प्रभावित होता है

✓ इसी कारण आधुनिक डिजिटल पुस्तकालयों में **ISBD + MARC + Metadata Standards** का **संयुक्त एवं पूरक प्रयोग** किया जाता है, जिससे संसाधनों की पहचान, खोज, साझाकरण और संरक्षण प्रभावी रूप से संभव हो सके।

8.12 व्यावहारिक उदाहरण (Integrated Example)

आधुनिक पुस्तकालयों में एक ही संसाधन के लिए विभिन्न मानकों का **संयुक्त (integrated)** प्रयोग किया जाता है।

Integrated Workflow

Printed Book → ISBD → MARC → Metadata → OPAC / Digital Library

व्याख्या :

मुद्रित पुस्तक का विवरण पहले ISBD के अनुसार मानकीकृत किया जाता है, फिर उसे MARC प्रारूप में मशीन-पठनीय बनाया जाता है और अंततः Metadata मानकों के माध्यम से OPAC एवं डिजिटल पुस्तकालयों में खोज व उपयोग के लिए उपलब्ध कराया जाता है।

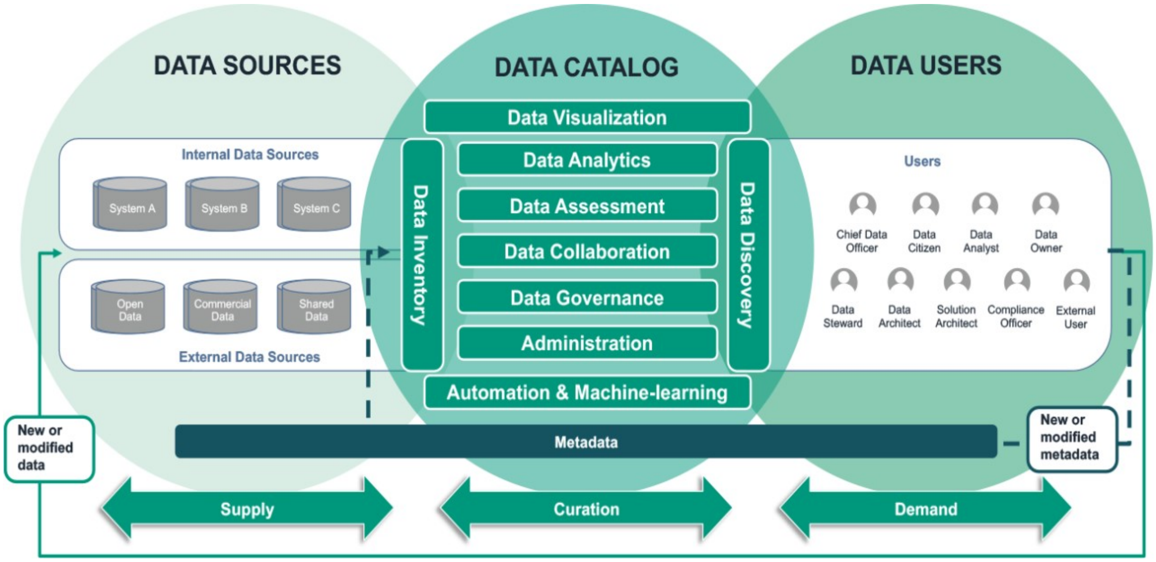
Imp.Point:

ISBD = Description | MARC = Encoding | Metadata = Discovery & Access

8.13 वर्तमान प्रवृत्तियाँ एवं भविष्य (Current Trends & Future)

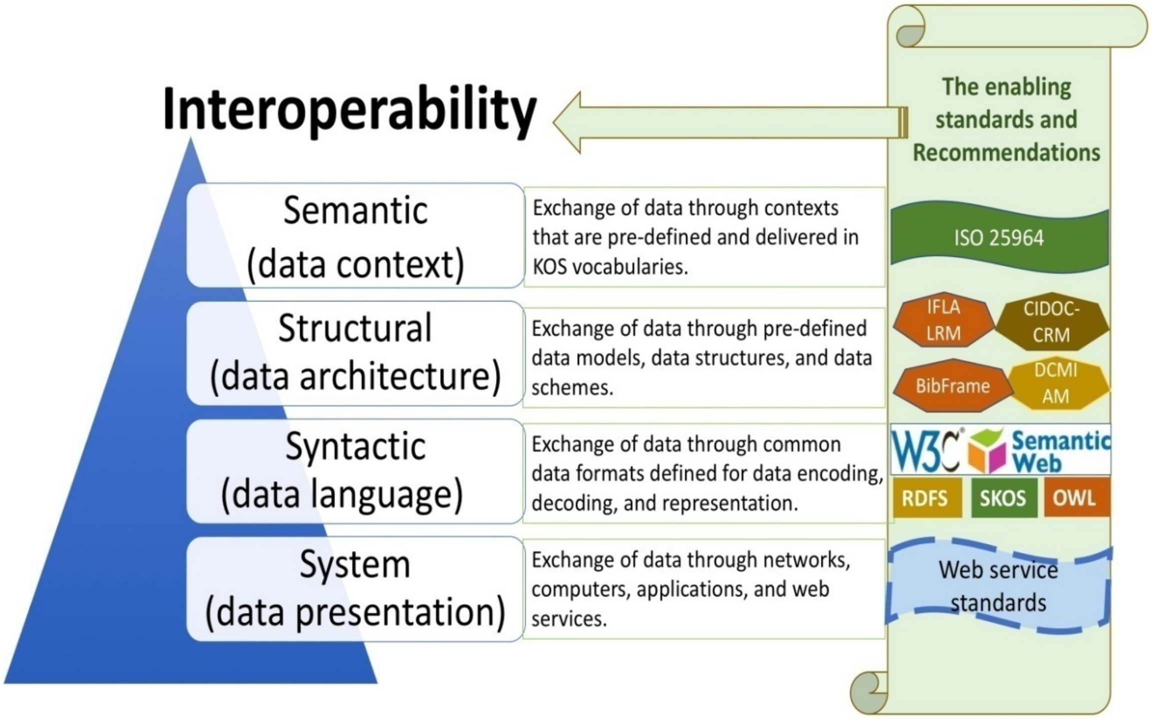
ग्रंथसूची विवरण एवं मानकीकरण के क्षेत्र में निम्नलिखित प्रमुख प्रवृत्तियाँ उभर रही हैं—

- **RDA-based Cataloguing (Resource Description and Access)** – आधुनिक, user-centric और digital-friendly cataloguing approach
- **Linked Data & Semantic Web** – ग्रंथसूची डेटा को web-of-data से जोड़कर **interoperability और discoverability** बढ़ाना
- **BIBFRAME (Bibliographic Framework)** – MARC का संभावित विकल्प, जो **Linked Data model** पर आधारित है



Future of cataloguing is shifting from MARC records to Linked Data models.

यह चित्र दर्शाता है कि आधुनिक cataloguing MARC से Linked Data एवं BIBFRAME जैसे web-based models की ओर बढ़ रही है, जिससे bibliographic data अधिक interoperable और globally discoverable बनता है।



यह चित्र दर्शाता है कि डिजिटल पुस्तकालयों में Interoperability केवल तकनीकी नहीं, बल्कि संरचनात्मक और अर्थगत स्तरों पर भी आवश्यक है। Linked Data और Semantic Web इन सभी स्तरों पर डेटा के प्रभावी आदान-प्रदान को संभव बनाते हैं।

8.14 सारांश

यह इकाई स्पष्ट करती है कि **ग्रंथसूची मानक (Bibliographic Standards)** आधुनिक पुस्तकालयों में **सूचना संगठन, खोज एवं साझाकरण** की आधारशिला हैं। **ISBD** ग्रंथसूची विवरण के तत्वों को एकरूप एवं मानकीकृत रूप प्रदान करता है, जिससे रिकॉर्ड अंतरराष्ट्रीय स्तर पर समझने योग्य बनते हैं। **MARC** इन विवरणों को **machine-readable format** में परिवर्तित करता है, जिससे पुस्तकालय automation और resource sharing संभव हो पाता है।
वहीं **Metadata standards** डिजिटल संसाधनों की पहचान, खोज, पहुँच तथा दीर्घकालिक संरक्षण को सुनिश्चित करते हैं और संसाधनों को **OPAC एवं डिजिटल पुस्तकालयों के माध्यम से वैश्विक मंच पर सुलभ** बनाते हैं। इस प्रकार ISBD, MARC और Metadata का संयुक्त प्रयोग आधुनिक डिजिटल पुस्तकालय सेवाओं को प्रभावी, interoperable और user-centric बनाता है।

8.15 शब्दावली (Glossary)

• **ISBD (International Standard Bibliographic Description)** – ग्रंथसूची विवरण के तत्वों के क्रम एवं निर्धारित विराम-चिह्न तय करने वाला अंतरराष्ट्रीय मानक।

• **MARC (Machine Readable Cataloguing)** – ग्रंथसूची रिकॉर्ड को कंप्यूटर-पठनीय बनाने वाला डेटा संचार प्रारूप।

• **Metadata** – "Data about Data"; किसी संसाधन के बारे में संरचित जानकारी।

• **Interoperability** –विभिन्न पुस्तकालय प्रणालियों के बीच डेटा साझा करने की क्षमता।

• **Dublin Core** – डिजिटल संसाधनों के वर्णन हेतु प्रयुक्त सरल और web-friendly metadata standard।

• **OPAC** –Online Public Access Catalogue, जिससे उपयोगकर्ता संसाधनों की खोज करता है।

• **RDA** –आधुनिक cataloguing standard (Resource Description and Access)।

• **Linked Data** – डेटा को वेब पर आपस में जोड़ने की तकनीक।

8.16 निबंधात्मक प्रश्न (Essay Questions)

प्रश्न 1. ग्रंथसूची मानकों की आवश्यकता स्पष्ट कीजिए।

उत्तर:
ग्रंथसूची मानकों की आवश्यकता इसलिए है ताकि पुस्तकालय संसाधनों का विवरण एकरूप, सटीक और साझा-योग्य बनाया जा सके। मानकों के अभाव में रिकॉर्ड असंगठित हो जाते हैं, जिससे खोज एवं resource sharing प्रभावित होता है। ISBD, MARC और Metadata जैसे मानक पुस्तकालय automation, interoperability और डिजिटल सूचना सेवाओं को प्रभावी बनाते हैं।

प्रश्न 2. ISBD और MARC के बीच अंतर स्पष्ट कीजिए।

उत्तर:
ISBD एक **descriptive standard** है, जो बताता है कि ग्रंथसूची विवरण किस क्रम और विराम-चिह्न के साथ लिखा जाए। MARC एक **encoding standard** है, जो उसी विवरण को machine-readable format में परिवर्तित करता है। अतः ISBD "क्या लिखना है" बताता है, जबकि MARC "कैसे encode करना है" बताता है।

प्रश्न 3. Metadata मानकों का डिजिटल पुस्तकालयों में महत्व बताइए।

उत्तर:
Metadata डिजिटल पुस्तकालयों में संसाधनों की पहचान, खोज, पहुँच और संरक्षण के लिए आवश्यक है। Dublin Core जैसे metadata standards डिजिटल संसाधनों को web-based discovery के लिए उपयुक्त बनाते हैं और global visibility बढ़ाते हैं।

प्रश्न 4. MARC 21 की प्रमुख विशेषताओं का वर्णन कीजिए।

उत्तर:
MARC 21 एक अंतरराष्ट्रीय रूप से प्रचलित machine-readable format है। इसमें Tag, Indicators और Subfields की संरचना होती है, जिससे ग्रंथसूची डेटा को कंप्यूटर द्वारा पढ़ा और साझा किया जा सकता है। अधिकांश आधुनिक पुस्तकालय MARC 21 का प्रयोग करते हैं।

प्रश्न 5. डिजिटल युग में ग्रंथसूची मानकों की भूमिका का विवेचन कीजिए।

उत्तर:
डिजिटल युग में ग्रंथसूची मानक सूचना संगठन की रीढ़ हैं। ISBD मानकीकरण प्रदान करता है, MARC automation को सक्षम बनाता है और Metadata डिजिटल संसाधनों को वैश्विक मंच पर सुलभ बनाता है। इनका संयुक्त प्रयोग आधुनिक पुस्तकालय सेवाओं को user-centric बनाता है।

8.17 लघु प्रश्न (Short Questions)

1. **MARC क्या है?**
   → MARC एक machine-readable cataloguing format है।
2. **Dublin Core क्या है?**
   → यह डिजिटल संसाधनों के लिए metadata standard है।
3. **Metadata क्यों आवश्यक है?**
   → संसाधनों की पहचान, खोज और पहुँच के लिए।
4. **Interoperability का अर्थ क्या है?**
   → विभिन्न प्रणालियों में डेटा साझा करने की क्षमता।
5. **ISBD का मुख्य उद्देश्य क्या है?**
   → ग्रंथसूची विवरण का मानकीकरण।

8.18 MCQs (उत्तर सहित)

1. ISBD किस संस्था द्वारा विकसित किया गया?
   (A) UNESCO
   (B) IFLA ✔
   (C) ALA
   (D) LC
2. MARC का मुख्य उद्देश्य है—
   (A) वर्गीकरण
   (B) मशीन-पठनीय रिकॉर्ड ✔
   (C) परिसंचरण
   (D) बजट
3. Dublin Core में कितने core elements होते हैं?
   (A) 10

(B) 12
(C) 15 ✔
(D) 20

4. Metadata का अर्थ है—
   (A) Data storage
   (B) Data about Data ✔
   (C) Data processing
   (D) Data security

5. MARC 21 किस देश द्वारा विकसित किया गया?
   (A) UK
   (B) India
   (C) USA ✔
   (D) Germany

6. OPAC का पूर्ण रूप है—
   (A) Online Public Access Catalogue ✔
   (B) Open Public Access Code
   (C) Online Processing Access Control
   (D) Open Programme Access Catalogue

7. Linked Data का संबंध किससे है?
   (A) Card Catalogue
   (B) Semantic Web ✔
   (C) Classification
   (D) Indexing

8. ISBN का प्रयोग होता है—
   (A) Journals के लिए
   (B) Books की पहचान के लिए ✔
   (C) Databases के लिए
   (D) Metadata harvesting के लिए

9. OAI-PMH का प्रयोग होता है—
   (A) Classification
   (B) Metadata harvesting ✔
   (C) Circulation
   (D) Cataloguing rules

10. BIBFRAME किसका संभावित विकल्प है?
   (A) ISBD
   (B) MARC ✔

(C) DDC
(D) AACR-2

8.19 संदर्भ ग्रंथ सूची (References):

1. आर. एल. शर्मा – *ग्रंथसूची संगठन के सिद्धांत*
2. एस. आर. रंगनाथन – Library Catalogue
3. यू. जी. सी. – ODL मार्गदर्शिका (MLIS)
4. IGNOU – Cataloguing and Metadata Study Material
5. बी. एन. शर्मा – डिजिटल पुस्तकालय
6. एस. के. सिंह – सूचना संगठन एवं पुनर्प्राप्ति
7. यू. के. मिश्र – आधुनिक पुस्तकालय विज्ञान
8. एन. के. गोयल – ग्रंथसूची नियंत्रण
9. के. एल. शर्मा – सूचना प्रबंधन
10. यूजीसी ई-पाठ्य सामग्री (MLIS)
11. IFLA. International Standard Bibliographic Description (ISBD)
12. Library of Congress. MARC 21 Format for Bibliographic Data
13. Dublin Core Metadata Initiative. Dublin Core Metadata Element Set
14. Taylor, A. G. Introduction to Cataloging and Classification
15. Hider, P. Information Resource Description
16. Svenonius, E. The Intellectual Foundation of Information Organization
17. Baca, M. Introduction to Metadata
18. Coyle, K. FRBR, Before and After
19. Miller, E. Metadata for Digital Libraries

तृतीय खण्ड

सूचना भंडारण और पुनर्प्राप्ति प्रणाली

(Information Storage and Retrieval Systems)

इकाई–9 : मेटाडाटा : MARC 21 (856 Field) एवं Dublin Core (Metadata: MARC 21–856 Field & Dublin Core)

इकाई की रूपरेखा

9.1 प्रस्तावना

9.2 उद्देश्य

9.3 मेटाडाटा: अर्थ एवं परिभाषा

9.4 मेटाडाटा के प्रकार

9.5 MARC 21 : परिचय एवं संरचना

9.6 MARC 21 के प्रमुख फील्ड्स का अवलोकन

9.7 MARC 21 – 856 Field

9.8 Dublin Core : अवधारणा एवं विकास

9.9 Dublin Core के 15 तत्त्व

9.10 MARC 21 एवं Dublin Core की तुलना

9.11 डिजिटल रिपॉजिटरी एवं वेब संसाधनों में मेटाडाटा

9.12 व्यावहारिक उदाहरण

9.13 मेटाडाटा की चुनौतियाँ

9.14 आधुनिक प्रवृत्तियाँ: Linked Data, BIBFRAME

9.15 सारांश

9.16 शब्दावली

9.17 निबंधात्मक प्रश्न

9.18 लघु प्रश्न एवं MCQ

9.19 अभ्यास प्रश्नों के उत्तर

9.20 संदर्भ ग्रंथ सूची

9.1 प्रस्तावना

डिजिटल युग में सूचना का स्वरूप तीव्र गति से परिवर्तित हुआ है। आज सूचना केवल मुद्रित पुस्तकों और पत्रिकाओं तक सीमित न होकर **ई-पुस्तकों, ई-जर्नलों, वेबसाइटों, डिजिटल रिपॉजिटरी तथा ऑनलाइन डेटाबेस** के रूप में व्यापक रूप से उपलब्ध है। इस डिजिटल वातावरण में सूचना संसाधनों की संख्या, प्रकार और स्वरूप में निरंतर वृद्धि हो रही है, जिसके परिणामस्वरूप सूचना का प्रभावी संगठन, नियंत्रण तथा उपयोग एक जटिल कार्य बन गया है।

डिजिटल सूचना संसाधनों की **खोज (Discovery), पहचान (Identification), पहुँच (Access)** और **प्रबंधन (Management)** तभी प्रभावी ढंग से संभव है, जब उनके साथ **मानकीकृत मेटाडाटा (Metadata)** संलग्न हो। मेटाडाटा किसी संसाधन के विषय-वस्तु, संरचना, तकनीकी विशेषताओं, अधिकार स्थिति तथा उपयोग संबंधी जानकारी प्रदान करता है। इस प्रकार मेटाडाटा डिजिटल सूचना संसाधनों और उपयोगकर्ताओं के बीच एक सेतु के रूप में कार्य करता है तथा सूचना पुनर्प्राप्ति (Information Retrieval) को अधिक सटीक और उपयोगी बनाता है।

डिजिटल संसाधनों के वर्णन और संगठन की आवश्यकता ने अंतरराष्ट्रीय स्तर पर विभिन्न **मेटाडाटा मानकों** के विकास को प्रेरित किया है। इनमें **MARC 21** और **Dublin Core** विशेष रूप से महत्वपूर्ण माने जाते हैं। MARC 21 पारंपरिक पुस्तकालय कैटलॉगिंग परंपरा पर आधारित एक विस्तृत और संरचित मेटाडाटा मानक है, जो डिजिटल संसाधनों के वर्णन के लिए भी व्यापक रूप से प्रयुक्त होता है। MARC 21 का **856 Field** इलेक्ट्रॉनिक संसाधनों के संदर्भ में विशेष महत्व रखता है, क्योंकि यह किसी संसाधन के इलेक्ट्रॉनिक स्थान, पहुँच विधि तथा डिजिटल संस्करण से संबंधित जानकारी प्रदान करता है।

दूसरी ओर, **Dublin Core** एक सरल, लचीला और वेब-उन्मुख मेटाडाटा मानक है, जिसे विशेष रूप से इंटरनेट आधारित सूचना संसाधनों के लिए विकसित किया गया है। इसके 15 मूल तत्त्व न्यूनतम लेकिन पर्याप्त विवरण प्रदान करते हैं, जिससे विभिन्न प्रणालियों और डिजिटल प्लेटफॉर्मों के बीच अंतर-संचालनीयता संभव हो पाती है। यही कारण है कि

डिजिटल रिपॉजिटरी, ओपन एक्सेस प्लेटफॉर्म तथा वेब संसाधनों में Dublin Core का व्यापक उपयोग किया जाता है।

अतः यह इकाई डिजिटल युग में मेटाडाटा की अवधारणा, उसके प्रकार, तथा **MARC 21 (विशेषतः 856 Field)** और **Dublin Core** जैसे अंतरराष्ट्रीय मानकों की भूमिका को समझने का सुदृढ़ आधार प्रदान करती है। यह इकाई विद्यार्थियों को डिजिटल सूचना संसाधनों के प्रभावी संगठन, खोज और प्रबंधन के लिए आवश्यक सैद्धान्तिक एवं व्यावहारिक ज्ञान प्रदान करने में सहायक सिद्ध होगी।

9.2 उद्देश्य

इस इकाई के अध्ययन के बाद विद्यार्थी—

- मेटाडाटा की **अवधारणा, प्रकार एवं भूमिका** समझ सकेंगे।
- **MARC 21** की संरचना तथा **856 Field** का व्यावहारिक उपयोग कर सकेंगे।
- **Dublin Core** के 15 तत्त्वों का विवरण एवं अनुप्रयोग समझेंगे।
- **MARC 21 बनाम Dublin Core** का तुलनात्मक विश्लेषण कर सकेंगे।
- डिजिटल रिपॉजिटरी/वेब संसाधनों के लिए **उपयुक्त मेटाडाटा स्कीमा** चुन सकेंगे।

9.3 मेटाडाटा : अर्थ एवं परिभाषा (Metadata: Meaning and Definition)

मेटाडाटा (Metadata) का शाब्दिक एवं सरल अर्थ है—
"डेटा के बारे में डेटा (Data about Data)"

अर्थात्, मेटाडाटा वह सूचना है जो किसी अन्य सूचना संसाधन के बारे में अतिरिक्त और व्याख्यात्मक जानकारी प्रदान करती है। यह जानकारी संसाधन की पहचान, समझ, खोज, उपयोग तथा प्रबंधन को सरल और प्रभावी बनाती है।

परिभाषा

मेटाडाटा वह **संरचित (structured) और मानकीकृत सूचना** है, जो किसी संसाधन—चाहे वह **भौतिक (Printed)** हो या **डिजिटल (Electronic)**—के **विवरण, संरचना, तकनीकी विशेषताओं, प्रबंधन तथा अधिकार (Rights)** से संबंधित जानकारी प्रदान करती है। डिजिटल वातावरण में मेटाडाटा सूचना संसाधनों को सुव्यवस्थित रूप से संगठित करने तथा उन्हें खोज योग्य (discoverable) बनाने का आधार प्रदान करती है।

दूसरे शब्दों में, मेटाडाटा किसी संसाधन की केवल पहचान ही नहीं कराती, बल्कि यह यह भी स्पष्ट करती है कि वह संसाधन **क्या है, किसके द्वारा निर्मित है, कैसे उपयोग किया जा सकता है, और किन शर्तों के अंतर्गत उपलब्ध है।**

उदाहरण

- किसी **ई-पुस्तक (E-Book)** से संबंधित निम्नलिखित सूचनाएँ मेटाडाटा के अंतर्गत आती हैं—
    - शीर्षक (Title)
    - लेखक (Author/Creator)
    - URL या Identifier
    - फ़ाइल-प्रकार (PDF, EPUB आदि)
    - भाषा (Language)
    - लाइसेंस एवं उपयोग अधिकार (Rights)
- पारंपरिक पुस्तकालयों में प्रयुक्त **कार्ड-कैटलॉग** मेटाडाटा का एक उत्कृष्ट पारंपरिक उदाहरण है, जिसमें पुस्तक का शीर्षक, लेखक, विषय, प्रकाशन विवरण आदि सूचनाएँ सम्मिलित होती हैं। यह कार्ड-कैटलॉग मुद्रित संसाधनों की खोज और पहचान में वही भूमिका निभाता है, जो डिजिटल वातावरण में मेटाडाटा निभाता है।

शैक्षणिक महत्व (Academic Relevance)

डिजिटल युग में सूचना संसाधनों की अत्यधिक वृद्धि के कारण मेटाडाटा का महत्व और भी बढ़ गया है। मेटाडाटा के बिना डिजिटल संसाधन अव्यवस्थित हो जाते हैं और उनकी प्रभावी खोज एवं पुनर्प्राप्ति संभव नहीं हो पाती। इसलिए मेटाडाटा को डिजिटल पुस्तकालयों, ऑनलाइन डेटाबेस तथा रिपॉजिटरी प्रणालियों की **आधारशिला (Foundation)** माना जाता है।

9.4 मेटाडाटा के प्रकार (Types of Metadata)

मेटाडाटा को उसके **कार्य, उद्देश्य और उपयोग** के आधार पर मुख्यतः **तीन प्रमुख प्रकारों** में वर्गीकृत किया जाता है। यह वर्गीकरण डिजिटल संसाधनों के प्रभावी संगठन, खोज और प्रबंधन को समझने में सहायक होता है। प्रत्येक प्रकार का मेटाडाटा सूचना संसाधन के किसी विशिष्ट पक्ष को स्पष्ट करता है।

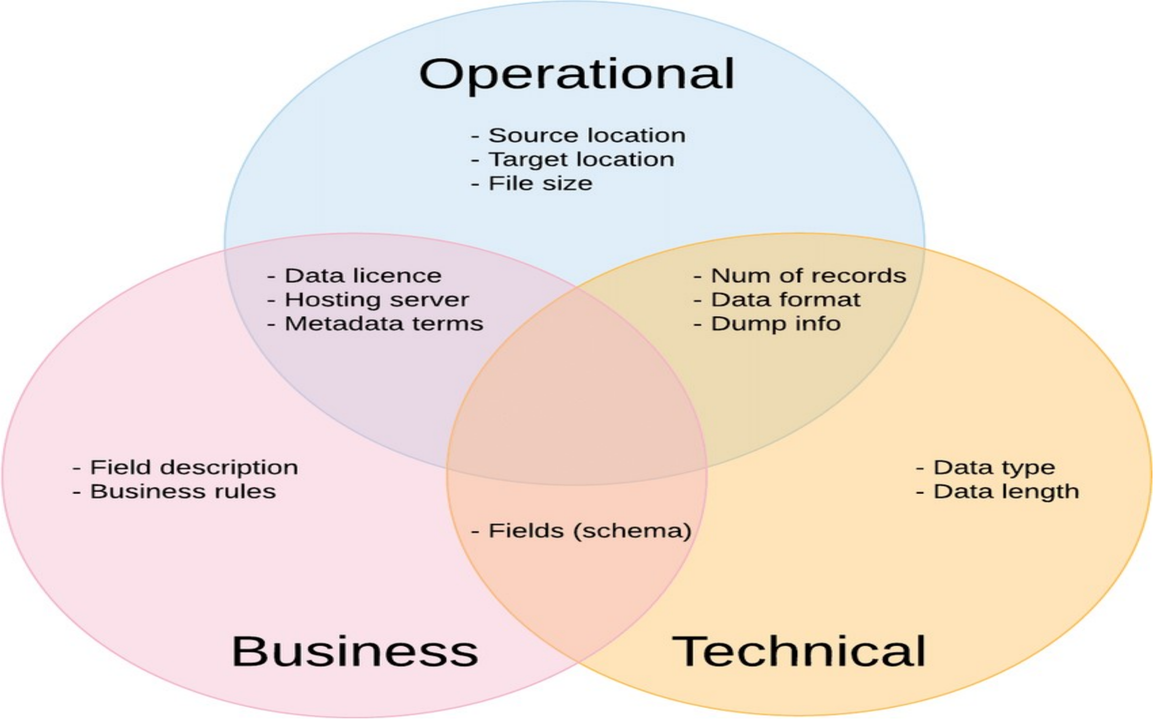
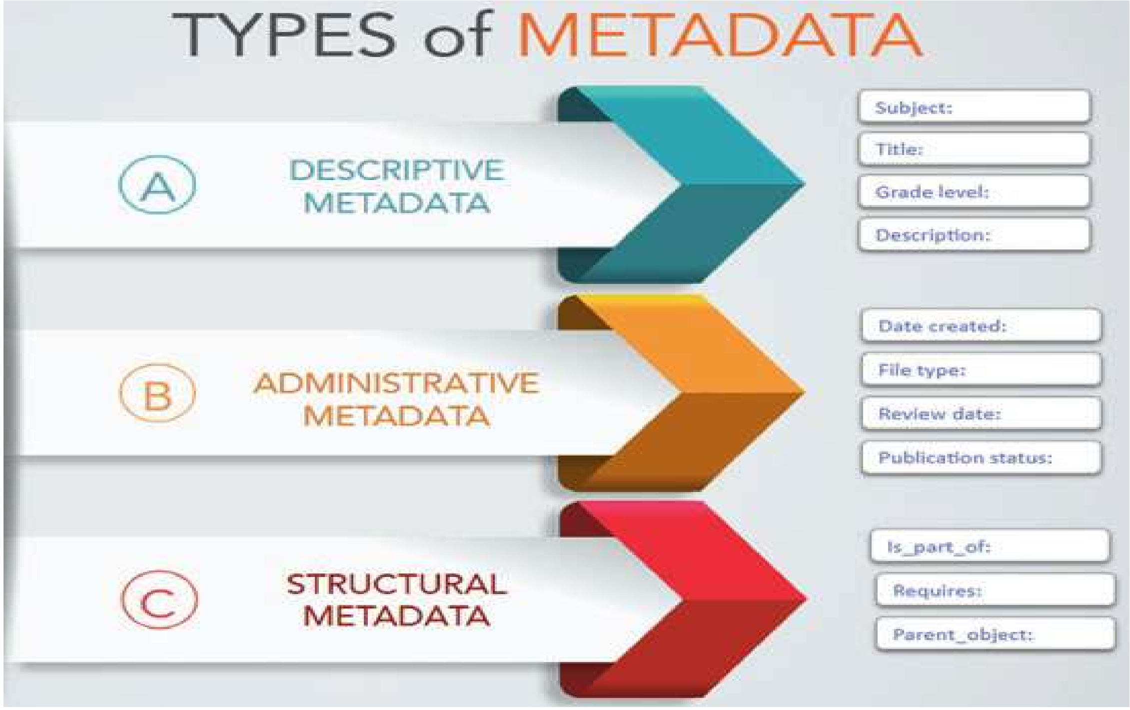


Fig.: Types of Metadata



9.4.1 वर्णनात्मक मेटाडाटा (Descriptive Metadata)

वर्णनात्मक मेटाडाटा का मुख्य उद्देश्य किसी सूचना संसाधन की **पहचान (Identification)** और **खोज (Discovery)** को सरल बनाना होता है। यह मेटाडाटा उपयोगकर्ता को यह समझने में सहायता करता है कि कोई संसाधन **क्या है, किस विषय से संबंधित है और किसके द्वारा निर्मित किया गया है।**

इस प्रकार का मेटाडाटा प्रायः **सूचना पुनर्प्राप्ति प्रणालियों (Information Retrieval Systems)**, OPAC तथा सर्च इंजनों में संसाधनों की खोज के लिए प्रयुक्त होता है।

प्रमुख तत्त्व (Elements):

- शीर्षक (Title)
- लेखक/निर्माता (Author / Creator)
- विषय (Subject)
- कीवर्ड्स (Keywords)
- सार/विवरण (Abstract / Description)

उदाहरण:
यदि कोई उपयोगकर्ता "Digital Libraries" विषय पर ई-पुस्तक खोजता है, तो वही ई-पुस्तक खोज परिणामों में दिखाई देगी, जिसका वर्णनात्मक मेटाडाटा उस विषय से संबद्ध होगा।

9.4.2 संरचनात्मक मेटाडाटा (Structural Metadata)

संरचनात्मक मेटाडाटा किसी सूचना संसाधन की **आंतरिक संरचना (Internal Structure)** तथा उसके विभिन्न भागों के **आपसी संबंधों** को स्पष्ट करता है। यह मेटाडाटा यह बताता है कि संसाधन के विभिन्न घटक किस प्रकार परस्पर जुड़े हुए हैं।

डिजिटल संसाधनों, विशेष रूप से **ई-पुस्तकों, डिजिटल दस्तावेजों और मल्टीमीडिया फाइलों** में संरचनात्मक मेटाडाटा अत्यंत महत्वपूर्ण भूमिका निभाता है।

प्रमुख उदाहरण:

- सामग्री-सूची (Table of Contents – TOC)
- अध्याय → अनुभाग → पृष्ठ (Chapter → Section → Page)
- एक डिजिटल दस्तावेज़ में टेक्स्ट, चित्र और तालिकाओं के बीच संबंध

उपयोगिता:
संरचनात्मक मेटाडाटा के माध्यम से उपयोगकर्ता सीधे किसी विशेष अध्याय, अनुभाग या पृष्ठ तक पहुँच सकता है, जिससे संसाधन का उपयोग अधिक सुविधाजनक हो जाता है।

9.4.3 प्रशासनिक मेटाडाटा (Administrative Metadata)

प्रशासनिक मेटाडाटा सूचना संसाधनों के **प्रबंधन (Management)**, **तकनीकी नियंत्रण (Technical Control)** तथा **अधिकार (Rights Management)** से संबंधित जानकारी प्रदान करता है। यह मेटाडाटा डिजिटल संसाधनों के दीर्घकालिक संरक्षण और सुरक्षित उपयोग के लिए आवश्यक होता है।

इस प्रकार का मेटाडाटा मुख्यतः **पुस्तकालय प्रशासकों, सिस्टम मैनेजरों और डिजिटल रिपॉजिटरी प्रबंधकों** के लिए उपयोगी होता है।

प्रमुख तत्त्व:

- फ़ाइल प्रारूप (File Format – PDF, JPEG, MP4 आदि)
- निर्माण तिथि (Creation Date)
- संशोधन तिथि (Modification Date)
- एक्सेस अधिकार (Access Rights)
- लाइसेंस एवं कॉपीराइट सूचना (License / Copyright)

उदाहरण:
किसी ई-जर्नल लेख पर "Open Access" या "Restricted Access" जैसी जानकारी प्रशासनिक मेटाडाटा का भाग होती है, जो यह निर्धारित करती है कि उस संसाधन का उपयोग कौन और कैसे कर सकता है।

शैक्षणिक दृष्टि से महत्त्व (Academic Significance)

तीनों प्रकार के मेटाडाटा—**वर्णनात्मक, संरचनात्मक और प्रशासनिक**—मिलकर डिजिटल सूचना संसाधनों के **समग्र संगठन और प्रभावी उपयोग** को संभव बनाते हैं। किसी भी डिजिटल पुस्तकालय या रिपॉजिटरी प्रणाली की सफलता इस बात पर निर्भर करती है कि इन तीनों प्रकार के मेटाडाटा का संतुलित और मानकीकृत उपयोग किस हद तक किया गया है।

9.5 MARC 21 : परिचय एवं संरचना (MARC 21: Introduction and Structure):

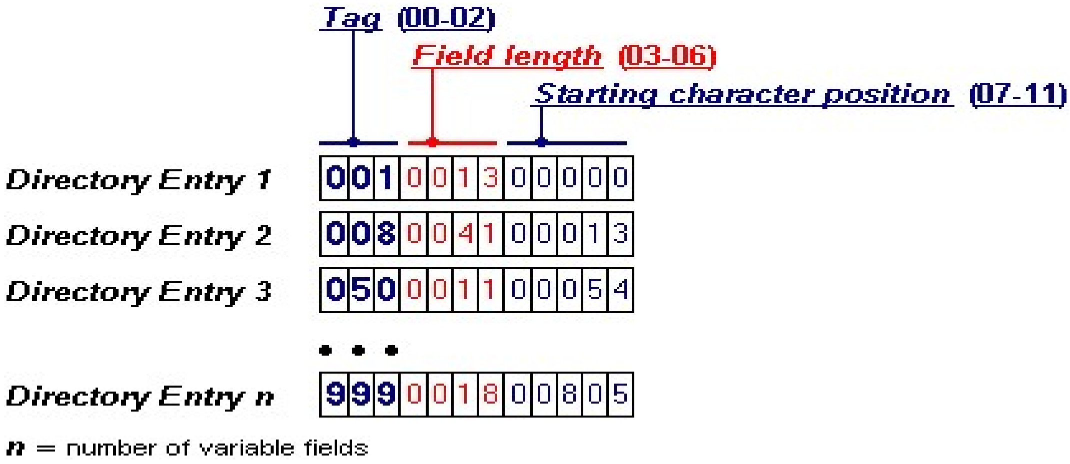
MARC 21 (Machine-Readable Cataloging) एक अंतरराष्ट्रीय स्तर पर स्वीकृत मेटाडाटा मानक है, जिसका उपयोग पुस्तकालयों और सूचना केंद्रों में **बिब्लियोग्राफिक रिकॉर्ड को मशीन-पठनीय (machine-readable) रूप** में प्रस्तुत करने के लिए किया जाता है। MARC 21 का मुख्य उद्देश्य यह सुनिश्चित करना है कि बिब्लियोग्राफिक सूचना को कंप्यूटर प्रणालियों द्वारा **संग्रहित, संसाधित, आदान-प्रदान (exchange)** तथा **प्रदर्शित** किया जा सके।

पारंपरिक कैटलॉगिंग में जहाँ बिब्लियोग्राफिक विवरण मानव-पठनीय रूप में होता था, वहीं MARC 21 उस विवरण को **संरचित कोड, टैग, इंडिकेटर और सब-फील्ड्स** के माध्यम से इस प्रकार प्रस्तुत करता है कि कंप्यूटर उसे आसानी से समझ और उपयोग कर सके। इस कारण MARC 21 को **पुस्तकालय स्वचालन (Library Automation)** और **OPAC प्रणालियों** की आधारशिला माना जाता है।

MARC 21 का विकास एवं रख-रखाव **Library of Congress** द्वारा किया जाता है, तथा यह मानक विश्वभर के पुस्तकालयों में बिब्लियोग्राफिक डेटा के आदान-प्रदान के लिए व्यापक रूप से अपनाया गया है।

MARC 21 रिकॉर्ड की संरचना (Structure of MARC 21 Record)

एक MARC 21 रिकॉर्ड मुख्यतः **तीन संरचनात्मक भागों** में विभाजित होता है। ये तीनों भाग मिलकर बिब्लियोग्राफिक सूचना को एक सुव्यवस्थित और मशीन-पठनीय रूप प्रदान करते हैं



1. Leader (लीडर)

Leader, MARC 21 रिकॉर्ड का पहला और अनिवार्य भाग होता है। इसकी लंबाई **24 characters** की निश्चित होती है। Leader में रिकॉर्ड से संबंधित **कोडित जानकारी (coded data)** होती है, जो यह निर्धारित करती है कि रिकॉर्ड को कंप्यूटर प्रणाली किस प्रकार संसाधित (process) करेगी।

Leader के प्रमुख कार्य:

- रिकॉर्ड का प्रकार (Record type)
- बिब्लियोग्राफिक स्तर (Bibliographic level)
- रिकॉर्ड की प्रोसेसिंग आवश्यकताएँ

Leader स्वयं मानव-पठनीय विवरण प्रदान नहीं करता, बल्कि यह रिकॉर्ड की **तकनीकी पहचान और नियंत्रण** का कार्य करता है।

2. Directory (डायरेक्टरी)

Directory, Leader के तुरंत बाद स्थित होती है। इसमें रिकॉर्ड के सभी **Variable Fields** की सूची होती है। प्रत्येक प्रविष्टि में यह जानकारी शामिल होती है कि—

- कौन-सा फील्ड किस **Tag** से पहचाना गया है
- वह फील्ड रिकॉर्ड में **कहाँ से शुरू** होता है
- उसकी **लंबाई (Length)** कितनी है

Directory का मुख्य उद्देश्य यह सुनिश्चित करना है कि कंप्यूटर प्रणाली रिकॉर्ड के विभिन्न फील्ड्स तक **तेज़ और सटीक पहुँच** प्राप्त कर सके।

3. Variable Fields (परिवर्ती क्षेत्र)

Variable Fields MARC 21 रिकॉर्ड का सबसे महत्वपूर्ण भाग होते हैं, क्योंकि इनमें वास्तविक **बिब्लियोग्राफिक डेटा** संग्रहित रहता है। इनकी लंबाई निश्चित नहीं होती और आवश्यकता के अनुसार बदल सकती है।

Variable Fields को दो भागों में विभाजित किया जाता है—

(a) Control Fields

- सामान्यतः **00X** टैग में होते हैं
- इनमें इंडिकेटर या सब-फील्ड नहीं होते

- उदाहरण: 001 (Control Number), 008 (Fixed Data Elements)

(b) Data Fields

- इनमें **Tag, Indicators** और **Subfields** होते हैं
- उदाहरण:
    - 100 – Author
    - 245 – Title
    - 260/264 – Publication
    - **856 – Electronic Location and Access**

Variable Fields के माध्यम से ही MARC 21 डिजिटल तथा मुद्रित दोनों प्रकार के संसाधनों का विस्तृत वर्णन संभव बनाता है।

9.6 MARC 21 के प्रमुख फील्ड्स

MARC 21 बिब्लियोग्राफिक रिकॉर्ड अनेक फील्ड्स से मिलकर बनता है, जिनमें प्रत्येक फील्ड संसाधन की किसी विशिष्ट सूचना को प्रदर्शित करता है। MARC 21 के कुछ प्रमुख फील्ड्स निम्नलिखित हैं—

- **100 – मुख्य प्रविष्टि (Main Entry–Author)**
  यह फील्ड संसाधन के **मुख्य लेखक या व्यक्तिगत उत्तरदायित्व** को दर्शाता है। यह फील्ड लेखक के नाम के आधार पर खोज एवं पहचान में सहायक होती है।
- **245 – शीर्षक फील्ड (Title Statement)**
  इस फील्ड में संसाधन का **मुख्य शीर्षक, उपशीर्षक तथा उत्तरदायित्व कथन** (statement of responsibility) अंकित किया जाता है। यह MARC 21 का सबसे महत्वपूर्ण वर्णनात्मक फील्ड माना जाता है।
- **260 / 264 – प्रकाशन विवरण (Publication Information)**
  इन फील्ड्स के माध्यम से संसाधन का **प्रकाशन स्थान, प्रकाशक का नाम तथा प्रकाशन तिथि** प्रदर्शित की जाती है। आधुनिक कैटलॉगिंग में 264 फील्ड का अधिक प्रचलन है।
- **300 – भौतिक विवरण (Physical Description)**
  यह फील्ड संसाधन के **आकार, पृष्ठ संख्या, चित्रों की उपस्थिति तथा अन्य भौतिक विशेषताओं** की जानकारी प्रदान करती है।

- **520 – सार या संक्षेप (Summary / Abstract)**
  इस फील्ड में संसाधन की **विषय-वस्तु का संक्षिप्त विवरण** दिया जाता है, जिससे उपयोगकर्ता यह समझ सके कि संसाधन उसके लिए प्रासंगिक है या नहीं।
- **650 – विषय प्रविष्टि (Subject Added Entry)**
  यह फील्ड संसाधन के **विषय शीर्षकों** को प्रदर्शित करती है और विषय आधारित खोज (subject searching) में अत्यंत सहायक होती है।
- **856 – इलेक्ट्रॉनिक स्थान एवं पहुँच (Electronic Location and Access)**
  यह फील्ड डिजिटल संसाधनों के संदर्भ में विशेष महत्व रखती है, क्योंकि इसके माध्यम से संसाधन के **URL, ऑनलाइन लोकेशन तथा इलेक्ट्रॉनिक एक्सेस** की जानकारी प्रदान की जाती है।

9.7 MARC 21 – 856 Field

डिजिटल सूचना संसाधनों के बढ़ते उपयोग के साथ MARC 21 में **इलेक्ट्रॉनिक संसाधनों की लोकेशन एवं पहुँच** को दर्शाने के लिए **856 Field** को विशेष रूप से परिभाषित किया गया है। यह फील्ड बिब्लियोग्राफिक रिकॉर्ड को ऑनलाइन संसाधनों से जोड़ने का कार्य करती है और उपयोगकर्ता को सीधे डिजिटल सामग्री तक पहुँच प्रदान करती है।

9.7.1 उद्देश्य (Purpose of 856 Field)

MARC 21 का **856 Field** किसी सूचना संसाधन के **इलेक्ट्रॉनिक स्थान (URL/URI)** तथा उसकी **पहुंच विधि (Access Method)** से संबंधित जानकारी प्रदान करने के लिए प्रयुक्त होता है।
इस फील्ड के माध्यम से यह स्पष्ट किया जाता है कि—

- कोई संसाधन ऑनलाइन उपलब्ध है या नहीं,
- वह किस माध्यम (जैसे HTTP, FTP आदि) से सुलभ है, तथा
- डिजिटल संसाधन का मूल बिब्लियोग्राफिक रिकॉर्ड से क्या संबंध है।

इस प्रकार 856 Field डिजिटल पुस्तकालयों, OPAC और ऑनलाइन कैटलॉग में **प्रत्यक्ष एक्सेस (Direct Access)** को संभव बनाती है।

9.7.2 Indicators

856 Field में **दो Indicators** होते हैं, जो इलेक्ट्रॉनिक संसाधन की पहुँच और उसके संबंध को स्पष्ट करते हैं।

(क) First Indicator – Access Method

यह Indicator संसाधन तक पहुँच की विधि को दर्शाता है—

- **0** – Email
- **1** – FTP
- **2** – Telnet
- **3** – Dial-up
- **4** – HTTP
- **7** – अन्य विधि (जिसे subfield **$2** में निर्दिष्ट किया जाता है)

(ख) Second Indicator – Relationship

यह Indicator इलेक्ट्रॉनिक संसाधन और रिकॉर्ड में वर्णित संसाधन के बीच संबंध को दर्शाता है—

- **0** – मूल संसाधन (Resource)
- **1** – संसाधन का संस्करण (Version of resource)
- **2** – संबंधित संसाधन (Related resource)

9.7.3 प्रमुख Subfields (Important Subfields)

856 Field में अनेक Subfields होते हैं, जिनमें से कुछ प्रमुख Subfields निम्नलिखित हैं—

- **$u – Uniform Resource Locator (URL)**
  इसमें संसाधन का वास्तविक वेब पता (URL/URI) दिया जाता है।
- **$z – Public Note**
  उपयोगकर्ताओं के लिए सार्वजनिक टिप्पणी या सूचना प्रदान करता है, जैसे एक्सेस की स्थिति।
- **$3 – Materials Specified**
  यह बताता है कि URL संसाधन के किस भाग से संबंधित है, जैसे *Table of Contents* या *Full Text*।
- **$2 – Access Method**
  तब प्रयुक्त किया जाता है, जब First Indicator का मान **7** हो और एक्सेस विधि अलग से बतानी हो।

उदाहरण :

856 4# $u https://example.org/ebook.pdf $z Full text available online

व्याख्या:
इस उदाहरण में संसाधन HTTP के माध्यम से उपलब्ध है और उपयोगकर्ता को पूर्ण पाठ (Full Text) तक ऑनलाइन पहुँच प्राप्त होती है।

9.8 Dublin Core : अवधारणा एवं विकास

Dublin Core Metadata Initiative (DCMI) एक सरल, लचीला और वेब-अनुकूल मेटाडाटा मानक है, जिसे **डिजिटल एवं वेब-आधारित सूचना संसाधनों** के प्रभावी वर्णन और खोज के लिए विकसित किया गया है। इसका उद्देश्य न्यूनतम लेकिन पर्याप्त मेटाडाटा प्रदान करना है, जिससे विभिन्न डिजिटल प्रणालियों के बीच **अंतर-संचालनीयता** संभव हो सके।

Dublin Core का विकास **1995 में Dublin (Ohio, USA)** में हुआ। इसका प्रमुख लक्ष्य **वेब संसाधनों की खोज-क्षमता (Resource Discovery)** को बेहतर बनाना है। इसी कारण इसका व्यापक उपयोग डिजिटल रिपॉजिटरी, ओपन एक्सेस प्लेटफॉर्म और संस्थागत संग्रहों में किया जाता है।

9.9 Dublin Core के 15 तत्त्व (The 15 Elements of Dublin Core Metadata):

Dublin Core में कुल **15 मूल तत्त्व (Core Elements)** होते हैं, जो किसी डिजिटल या वेब-आधारित सूचना संसाधन का **मानकीकृत, संक्षिप्त एवं प्रभावी विवरण** प्रदान करते हैं। ये तत्त्व संसाधनों की **खोज, पहचान, उपयोग और प्रबंधन** में सहायक होते हैं।

1. **Title (शीर्षक)**
   संसाधन का नाम या शीर्षक, जिससे उसकी प्राथमिक पहचान होती है।
2. **Creator (निर्माता)**
   वह व्यक्ति या संस्था जिसने संसाधन का निर्माण किया हो।
3. **Subject (विषय)**
   संसाधन का मुख्य विषय या विषय-वस्तु, सामान्यतः कीवड्र्स या विषय शीर्षकों के रूप में।
4. **Description (विवरण)**
   संसाधन की विषय-वस्तु का संक्षिप्त विवरण या सार।
5. **Publisher (प्रकाशक)**
   वह संस्था या निकाय जो संसाधन को प्रकाशित या उपलब्ध कराता है।

6. **Contributor (योगदानकर्ता)**
   वे व्यक्ति या संस्थाएँ जिन्होंने संसाधन के निर्माण में सहयोग किया हो।
7. **Date (तिथि)**
   संसाधन से संबंधित महत्वपूर्ण तिथि, जैसे निर्माण या प्रकाशन तिथि।
8. **Type (प्रकार)**
   संसाधन की प्रकृति या श्रेणी, जैसे Text, Image, Audio आदि।
9. **Format (स्वरूप)**
   संसाधन का तकनीकी प्रारूप, जैसे PDF, HTML, JPEG आदि।
10. **Identifier (पहचानकर्ता)**
    संसाधन की विशिष्ट पहचान, जैसे URL, URI या DOI।
11. **Source (स्रोत)**
    वह मूल संसाधन जिससे यह संसाधन व्युत्पन्न हुआ हो।
12. **Language (भाषा)**
    वह भाषा जिसमें संसाधन की विषय-वस्तु उपलब्ध है।
13. **Relation (संबंध)**
    संसाधन का अन्य संबंधित संसाधनों से संबंध।
14. **Coverage (परिसर)**
    संसाधन का भौगोलिक या कालिक क्षेत्र।
15. **Rights (अधिकार)**
    संसाधन के उपयोग से संबंधित अधिकार, लाइसेंस या कॉपीराइट सूचना।

उदाहरण:

<dc:title>Digital Library Systems</dc:title>

<dc:creator>Ranganathan</dc:creator>

<dc:identifier>https://repo.edu/item/123</dc:identifier>

इस उदाहरण में—

- **dc:title** संसाधन के शीर्षक को दर्शाता है, जो यहाँ *Digital Library Systems* है।
- **dc:creator** संसाधन के निर्माता/लेखक को दर्शाता है, जो यहाँ *Ranganathan* हैं।

- **dc:identifier** संसाधन की विशिष्ट पहचान बताता है; यहाँ दिया गया URL उस संसाधन के ऑनलाइन उपलब्ध होने का पता दर्शाता है।

Element	Definition
Title	A name given to a resource
Creator	An entity primarily responsible for making the content of a resource
Subject	A topic of the content of a resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of a resource
Date	A data of an event in the lifecycle of a resource
Type	The nature or genre of the content of a resource
Format	The physical or digital format of a resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to an another resource from which a resource is derived
Language	A language of the content of a resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of a resource
Rights	Information about rights held in and over a re-source

Fig.: Dublin Core Metadata Elements

Explaination: यह तालिका किसी भी डिजिटल संसाधन (जैसे किताब, लेख, फ़ाइल) की पहचान और विवरण बताने के लिए उपयोग होने वाले **मानक मेटाडेटा तत्वों** की सूची दिखाती है।

Metadata Element	Purpose	Type
Title	General Title for Data-set	character string
Subject	Variable identifier	character string
Description	What data variable is collected?	character string
Publisher	Who posts it on the WEB?	character string
Contributor	Who collected/prepared it?	character string
Date	When the data collected?	real
Type	What type of data?	character string
Format	Is the data text/ASCII?	character string
Identifier	Pointer to the data file	URL
Source	Where did the data originate from?	character string
Language	What language is being used?	character string
Coverage	Temporal/spatial coverage of the data set?	
Box	What is the Geo-spatial extent?	
Northlimit	Northernmost limit	real
Southlimit	Southernmost limit	real
Westlimit	Westernmost limit	real
Eastlimit	Easternmost limit	real
Period	What is the temporal extent?	
Start	Time for first data digit [millisecs]	integer
End	Time for last data digit [millisecs]	integer
Frequency	Δt between one data item and the next	Real
Rights	Copyright identifier	character string

9.10 MARC 21 एवं Dublin Core की तुलना

MARC 21 और Dublin Core दोनों ही महत्वपूर्ण मेटाडाटा मानक हैं, किंतु इनके **उद्देश्य, संरचना और प्रयोग क्षेत्र** में स्पष्ट अंतर पाया जाता है। MARC 21 पारंपरिक पुस्तकालय कैटलॉगिंग आवश्यकताओं को ध्यान में रखकर विकसित किया गया है, जबकि Dublin Core वेब एवं डिजिटल वातावरण के लिए अधिक उपयुक्त है।

आधार	MARC 21	Dublin Core
जटिलता	अधिक जटिल एवं विस्तृत	सरल एवं लचीला
मुख्य उपयोग	पुस्तकालय कैटलॉग एवं OPAC	वेब संसाधन एवं डिजिटल रिपॉजिटरी
तत्त्वों की संख्या	सैकड़ों फील्ड्स एवं सब-फील्ड्स	15 मूल तत्त्व
तकनीकी दृष्टिकोण	लाइब्रेरी-केन्द्रित	वेब-केन्द्रित
संरचना	टैग, इंडिकेटर एवं सब-फील्ड आधारित	तत्त्व-आधारित
उपयोग-सुविधा	प्रशिक्षित कैटलॉगर हेतु	सामान्य उपयोगकर्ताओं हेतु

9.11 डिजिटल रिपॉजिटरी एवं वेब संसाधनों में मेटाडाटा

डिजिटल रिपॉजिटरी और वेब संसाधनों के संगठन एवं खोज के लिए मानकीकृत मेटाडाटा का उपयोग आवश्यक है। **DSpace** और **EPrints** जैसी रिपॉजिटरी प्रणालियों में सामान्यतः **Dublin Core** मेटाडाटा अपनाया जाता है, क्योंकि यह सरल, लचीला और वेब-आधारित संसाधनों के लिए उपयुक्त है।

वहीं **OPAC** तथा **Integrated Library Systems (ILS)** में **MARC 21** मेटाडाटा मानक का प्रयोग किया जाता है, जो विस्तृत बिब्लियोग्राफिक नियंत्रण प्रदान करता है।

इसके अतिरिक्त, **OAI-PMH** के माध्यम से विभिन्न डिजिटल रिपॉजिटरी से मेटाडाटा का harvesting किया जाता है, जिससे विभिन्न प्लेटफॉर्मों के बीच संसाधनों की खोज और साझा उपयोग संभव हो पाता है।

- **DSpace, EPrints** → Dublin Core
- **OPAC/ILS** → MARC 21
- **OAI-PMH** के माध्यम से मेटाडाटा harvesting

9.12 व्यावहारिक उदाहरण (Example): एक ई-जर्नल लेख

डिजिटल संसाधनों के विवरण के लिए **MARC 21** और **Dublin Core** दोनों का प्रयोग किया जाता है। निम्न उदाहरण एक ई-जर्नल लेख को सरल रूप में दर्शाता है।

MARC 21 (856 Field)

856 4# $u https://journal.org/article.pdf $z Open Access

अर्थ:
यह फील्ड लेख के **ऑनलाइन स्थान (URL)** को बताती है और यह भी स्पष्ट करती है कि लेख **Open Access** है।

Dublin Core

dc:title = Information Retrieval

dc:creator = Salton, G.

dc:identifier = https://journal.org/article.pdf

अर्थ: इस उदाहरण में `Dublin Core` के तत्त्वों द्वारा लेख का शीर्षक, लेखक तथा उसकी ऑनलाइन पहचान (`URL`) संक्षेप में प्रदर्शित की गई है।

9.13 मेटाडाटा की चुनौतियाँ (Challenges of Metadata)

डिजिटल सूचना संसाधनों के तीव्र विस्तार के साथ मेटाडाटा के उपयोग में अनेक व्यावहारिक एवं तकनीकी चुनौतियाँ उत्पन्न हुई हैं, जो इसके प्रभावी संगठन और प्रबंधन को प्रभावित करती हैं। प्रमुख चुनौतियाँ निम्नलिखित हैं—

- **असंगतता (Inconsistency)**
  विभिन्न संस्थानों और प्रणालियों द्वारा मेटाडाटा के अलग-अलग मानकों और प्रथाओं के प्रयोग से रिकॉर्ड्स में एकरूपता का अभाव पाया जाता है, जिससे खोज और साझा उपयोग में कठिनाई होती है।
- **बहुभाषिकता (Multilinguality)**
  वैश्विक डिजिटल वातावरण में मेटाडाटा अनेक भाषाओं में तैयार किया जाता है। इससे विषय-शीर्षकों, कीवर्ड्स और विवरणों के मानकीकरण में समस्या उत्पन्न होती है।
- **स्कीमा-मैपिंग (Schema Mapping)**
  विभिन्न मेटाडाटा स्कीमाओं (जैसे MARC 21 और Dublin Core) के बीच डेटा रूपांतरण और मैपिंग एक जटिल प्रक्रिया है, जिसमें सूचना की हानि या विकृति की संभावना रहती है।
- **दीर्घकालिक संरक्षण (Long-term Preservation)**
  तकनीकी परिवर्तनों के कारण मेटाडाटा का दीर्घकाल तक सुरक्षित और उपयोगी बने रहना एक बड़ी चुनौती है, विशेषकर डिजिटल रिपॉजिटरी के संदर्भ में।

9.14 आधुनिक प्रवृत्तियाँ (Linked Data एवं BIBFRAME)

डिजिटल युग में मेटाडाटा को अधिक **अर्थपूर्ण, जुड़ा हुआ और वेब-अनुकूल** बनाने के लिए कुछ आधुनिक प्रवृत्तियाँ विकसित हुई हैं।

1. Linked Data

Linked Data का उद्देश्य सूचना संसाधनों को **सेमान्टिक (अर्थ-आधारित) रूप से आपस में जोड़ना** है, ताकि विभिन्न डेटासेट्स के बीच स्पष्ट संबंध स्थापित हो सकें।

उदाहरण:

इस प्रकार सभी संसाधन आपस में जुड़े रहते हैं और खोज अधिक सटीक हो जाती है।

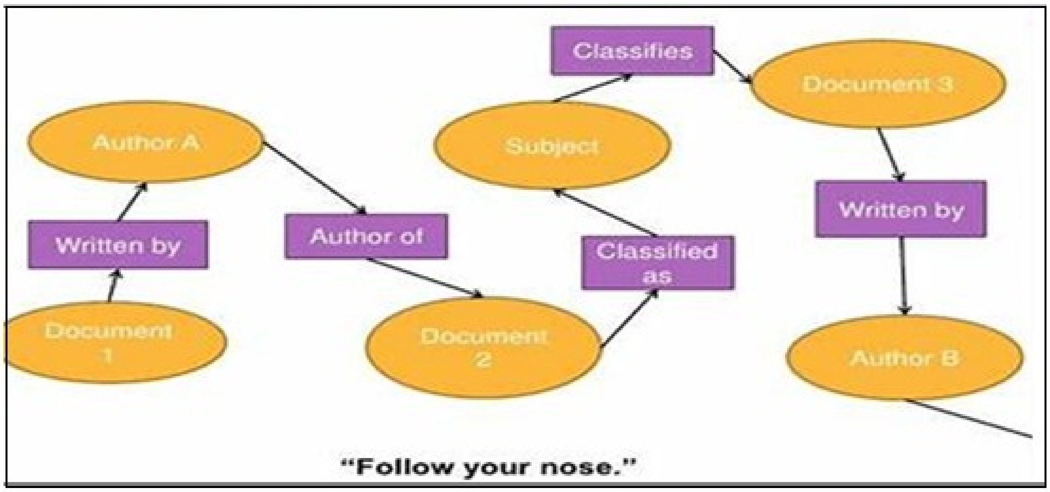


Fig.: Linked Data Frame

2. BIBFRAME

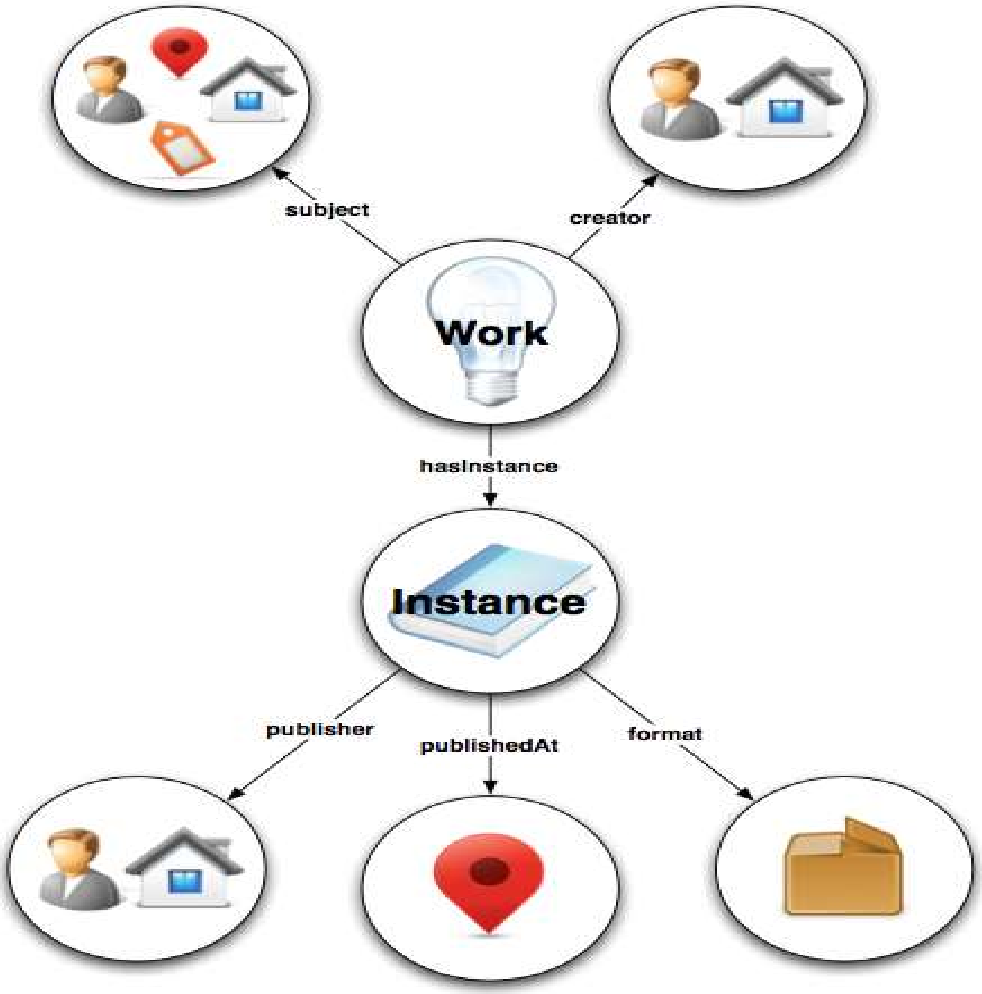
BIBFRAME (Bibliographic Framework), MARC का RDF-आधारित विकल्प है। इसका उद्देश्य पारंपरिक MARC रिकॉर्ड्स को **Linked Data और वेब वातावरण** के अनुरूप बनाना है।

सरल उदाहरण:

- **MARC रिकॉर्ड → BIBFRAME में**
    - ***Work* (कृति)**
    - ***Instance* (संस्करण)**

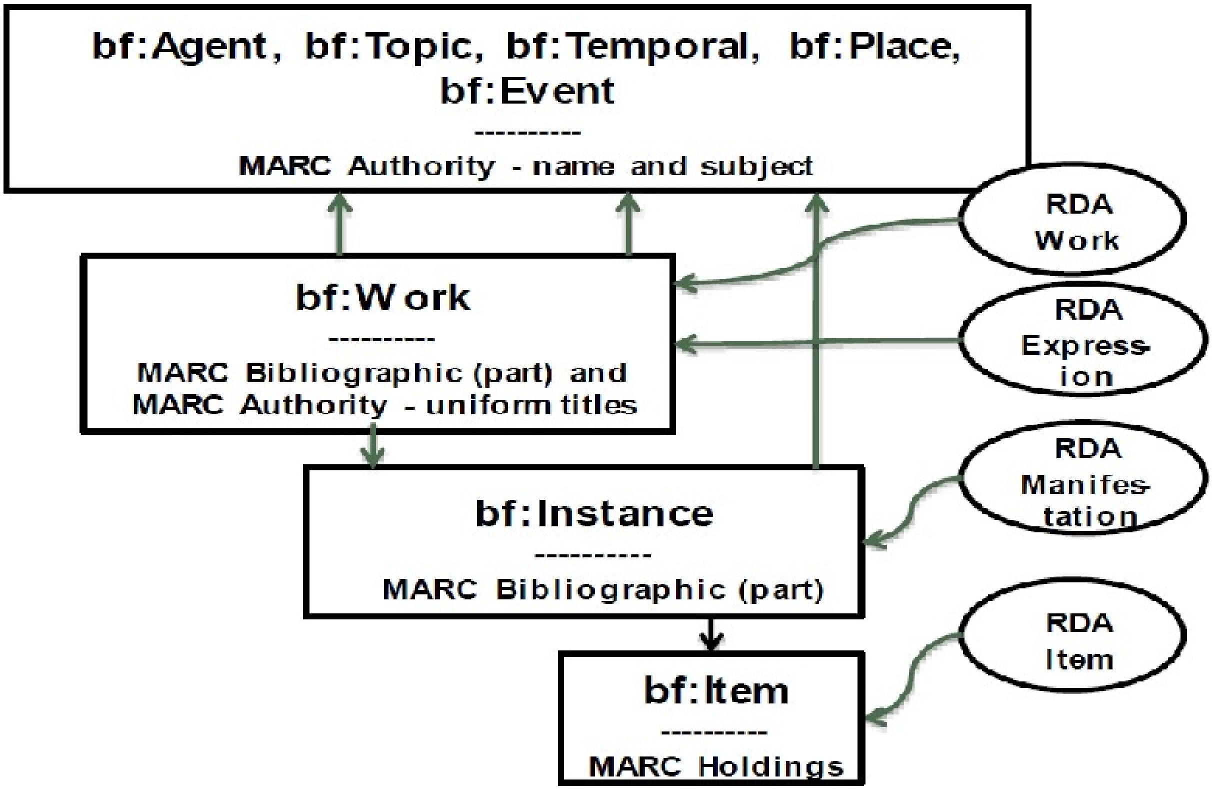
- ***Item*** (प्रति)

के रूप में संरचित किया जाता है।



BIBFRAME Work–Instance Model

Explainaition : यह चित्र दिखाता है कि **Work (कृति)** से उसका **Instance (संस्करण)** कैसे जुड़ा होता है और किस तरह creator, subject, publisher, format जैसी जानकारी जोड़ी जाती है।



Fg.: BIBFRAME–RDA–MARC Mapping Model

Explaination:
यह चित्र समझाता है कि **MARC रिकॉर्ड** को **BIBFRAME (Work–Instance–Item)** और **RDA मॉडल** से कैसे जोड़ा जाता है।

Linked Data संसाधनों को अर्थपूर्ण रूप से जोड़ता है, जबकि BIBFRAME बिब्लियोग्राफिक डेटा को आधुनिक वेब-आधारित ढाँचे में प्रस्तुत करता है।

9.14 आधुनिक प्रवृत्तियाँ (Modern Trends in Metadata)

डिजिटल सूचना वातावरण के निरंतर विकास के साथ मेटाडाटा के क्षेत्र में कुछ नई और महत्वपूर्ण प्रवृत्तियाँ उभरकर सामने आई हैं, जिनका उद्देश्य सूचना संसाधनों की **खोज, आपसी संबंध और अंतर-संचालनीयता** को और अधिक प्रभावी बनाना है।

- **Linked Data:** Linked Data का उद्देश्य विभिन्न सूचना संसाधनों को **सेमान्टिक (अर्थ-आधारित) रूप से आपस में जोड़ना** है। इसके माध्यम से अलग-अलग डेटासेट्स के बीच अर्थपूर्ण संबंध स्थापित किए जाते हैं, जिससे सूचना की खोज अधिक सटीक,

व्यापक और संदर्भ-संपन्न हो जाती है। यह तकनीक सेमान्टिक वेब की अवधारणा को सुदृढ़ बनाती है।

- **BIBFRAME (Bibliographic Framework): BIBFRAME** को MARC के आधुनिक विकल्प के रूप में विकसित किया गया है। यह **RDF (Resource Description Framework)** पर आधारित है और बिब्लियोग्राफिक डेटा को Linked Data वातावरण के अनुरूप प्रस्तुत करता है। BIBFRAME का उद्देश्य MARC रिकॉर्ड्स को वेब-आधारित, लचीले और भविष्य-उन्मुख ढाँचे में रूपांतरित करना है।

डिजिटल युग में मेटाडाटा के क्षेत्र में नई प्रवृत्तियाँ उभर रही हैं, जिनका उद्देश्य सूचना संसाधनों की खोज और आपसी संबंधों को बेहतर बनाना है।

9.15 सारांश

यह इकाई डिजिटल युग में मेटाडाटा के महत्व को स्पष्ट करती है और दर्शाती है कि **MARC 21 (विशेषतः 856 Field)** तथा **Dublin Core** डिजिटल संसाधनों की मानकीकृत पहचान, खोज और पहुँच के लिए आवश्यक मेटाडाटा मानक हैं। MARC 21 पुस्तकालय कैटलॉग में इलेक्ट्रॉनिक संसाधनों तक प्रत्यक्ष पहुँच प्रदान करता है, जबकि Dublin Core सरल और वेब-अनुकूल ढाँचे के माध्यम से डिजिटल रिपॉजिटरी एवं वेब संसाधनों के वर्णन में सहायक है। साथ ही, इकाई मेटाडाटा की चुनौतियों और **Linked Data** तथा **BIBFRAME** जैसी आधुनिक प्रवृत्तियों का संक्षिप्त परिचय भी प्रस्तुत करती है।

9.16 शब्दावली (Glossary)

- **Metadata (मेटाडाटा)**
  डेटा के बारे में डेटा; वह संरचित सूचना जो किसी संसाधन की पहचान, विवरण, खोज और प्रबंधन में सहायक होती है।
- **MARC 21 (Machine-Readable Cataloging)**
  एक अंतरराष्ट्रीय मेटाडाटा मानक, जिसका उपयोग बिब्लियोग्राफिक रिकॉर्ड को मशीन-पठनीय रूप में प्रस्तुत करने हेतु किया जाता है।
- **856 Field**
  MARC 21 का फील्ड जो डिजिटल संसाधन के इलेक्ट्रॉनिक स्थान (URL/URI) और ऑनलाइन पहुँच की जानकारी देता है।
- **Dublin Core**
  15 तत्त्वों वाला सरल और वेब-अनुकूल मेटाडाटा मानक, जो डिजिटल एवं वेब संसाधनों के वर्णन के लिए प्रयुक्त होता है।

- **DCMI (Dublin Core Metadata Initiative)**
  Dublin Core मेटाडाटा मानक के विकास और रख-रखाव हेतु उत्तरदायी अंतरराष्ट्रीय पहल।
- **URL (Uniform Resource Locator)**
  किसी डिजिटल संसाधन का विशिष्ट वेब पता, जिसके माध्यम से उसे इंटरनेट पर एक्सेस किया जाता है।
- **Digital Repository (डिजिटल रिपॉजिटरी)**
  वह प्रणाली जिसमें डिजिटल संसाधनों को मेटाडाटा सहित संग्रहीत और प्रबंधित किया जाता है।
- **Linked Data**
  संसाधनों को सेमान्टिक रूप से आपस में जोड़ने की तकनीक, जिससे सूचना की खोज अधिक प्रभावी बनती है।
- **BIBFRAME**
  MARC का RDF-आधारित आधुनिक विकल्प, जो बिब्लियोग्राफिक डेटा को वेब और Linked Data वातावरण के अनुकूल बनाता है।
- **OAI-PMH**
  डिजिटल रिपॉजिटरी से मेटाडाटा harvesting हेतु प्रयुक्त प्रोटोकॉल।

9.17 लघु प्रश्न एवं MCQ

A. लघु प्रश्न (Short Answer Questions)

1. मेटाडाटा को “Data about Data” क्यों कहा जाता है?
2. वर्णनात्मक (Descriptive) मेटाडाटा का मुख्य उद्देश्य क्या है?
3. MARC 21 में 856 Field का प्रयोग क्यों किया जाता है?
4. Dublin Core को वेब-अनुकूल मेटाडाटा स्कीमा क्यों कहा जाता है?
5. MARC 21 और Dublin Core में संरचनात्मक अंतर स्पष्ट कीजिए।
6. 856 Field में Subfield **$u** का महत्व क्या है?
7. Dublin Core में **Identifier** तत्त्व का प्रयोग क्यों आवश्यक है?
8. डिजिटल रिपॉजिटरी में मेटाडाटा की भूमिका स्पष्ट कीजिए।
9. प्रशासनिक मेटाडाटा के दो उदाहरण दीजिए।
10. Linked Data का मेटाडाटा से क्या संबंध है?

B. बहुविकल्पीय प्रश्न (MCQs)

1. मेटाडाटा का सबसे उपयुक्त अर्थ क्या है?
(a) डेटा का संग्रह
(b) सूचना का वर्गीकरण
(c) **डेटा के बारे में डेटा**
(d) डिजिटल संग्रह

✅उत्तर: (c)

2. निम्न में से कौन-सा वर्णनात्मक मेटाडाटा का उदाहरण है?
(a) File size
(b) Access rights
(c) **Title एवं Author**
(d) Compression method

✅उत्तर: (c)

3. MARC 21 का पूर्ण रूप क्या है?
(a) Machine Access Resource Code
(b) **Machine Readable Cataloging**
(c) Metadata Access Record Code
(d) Machine Resource Control

✅उत्तर: (b)

4. MARC 21 का कौन-सा फील्ड इलेक्ट्रॉनिक संसाधन की लोकेशन बताता है?
(a) 245
(b) 300
(c) 650
(d) **856**

✅उत्तर: (d)

5. 856 Field में Subfield $u किसके लिए प्रयुक्त होता है?
(a) Title
(b) Author
(c) **URL / URI**
(d) Access note

✅उत्तर: (c)

6. Dublin Core में कुल कितने मूल तत्त्व होते हैं?
(a) 10
(b) 12

(c) **15**
(d) 20

✅उत्तर: (c)

7. निम्न में से कौन-सा Dublin Core का तत्त्व नहीं है?
(a) Creator
(b) Subject
(c) **Call Number**
(d) Rights

✅उत्तर: (c)

8. Dublin Core Metadata Initiative (DCMI) का मुख्य उद्देश्य है—
(a) पुस्तकालय स्वचालन
(b) वर्गीकरण विकास
(c) **वेब संसाधनों की खोज में सुधार**
(d) मुद्रित ग्रंथ सूची निर्माण

✅उत्तर: (c)

9. MARC 21 की तुलना में Dublin Core किस प्रकार का स्कीमा है?
(a) अधिक जटिल
(b) कम लचीला
(c) **सरल एवं लचीला**
(d) केवल प्रिंट-आधारित

✅उत्तर: (c)

10. Linked Data का मुख्य लक्ष्य क्या है?
(a) डेटा का भंडारण
(b) डेटा का संपीड़न
(c) **डेटा के बीच अर्थपूर्ण संबंध स्थापित करना**
(d) डेटा का बैक-अप

✅उत्तर: (c)

11. MARC 21 में Second Indicator (856) का प्रयोग किसलिए होता है?
(a) Access method बताने के लिए

(b) URL की लंबाई दिखाने के लिए
(c) **संसाधन से संबंध बताने के लिए**
(d) फ़ाइल प्रकार दिखाने के लिए

✅उत्तर: (c)

12. Dublin Core का कौन-सा तत्त्व कानूनी उपयोग को दर्शाता है?
(a) Coverage
(b) Relation
(c) **Rights**
(d) Format

✅उत्तर: (c)

13. OAI-PMH का संबंध किससे है?
(a) Cataloguing rules
(b) **Metadata harvesting**
(c) Indexing language
(d) Authority control

✅उत्तर: (b)

14. BIBFRAME मुख्यतः किस तकनीक पर आधारित है?
(a) XML
(b) MARC
(c) **RDF / Linked Data**
(d) HTML

✅उत्तर: (c)

15. डिजिटल रिपॉजिटरी में Dublin Core अधिक प्रचलित होने का कारण है—
(a) कम फील्ड
(b) MARC से पुराना होना
(c) **सरलता और अंतर-संचालनीयता**
(d) केवल पुस्तकालयों तक सीमित होना

✅उत्तर: (c)

9.18 अभ्यास प्रश्नों के उत्तर

प्रश्न: मेटाडाटा क्या है?

उत्तर:
मेटाडाटा वह संरचित सूचना है जो किसी सूचना संसाधन की पहचान, संरचना, उपयोग, प्रबंधन और पहुँच से संबंधित विवरण प्रदान करती है। डिजिटल वातावरण में मेटाडाटा संसाधनों की खोज और पुनर्प्राप्ति को प्रभावी बनाता है।

प्रश्न: 856 Field का महत्व स्पष्ट कीजिए।

उत्तर:
MARC 21 का 856 Field इलेक्ट्रॉनिक संसाधनों के URL, एक्सेस विधि तथा संसाधन से उसके संबंध को दर्शाता है। यह डिजिटल लाइब्रेरी और OPAC में ऑनलाइन सामग्री तक प्रत्यक्ष पहुँच प्रदान करता है।

9.19 संदर्भ ग्रंथ सूची

1. शर्मा, आर. एन. पुस्तकालय स्वचालन एवं नेटवर्किंग, अटलांटा पब्लिशर्स, नई दिल्ली।
2. कुमार, पी. सूचना संगठन एवं मेटाडाटा, प्रभात प्रकाशन।
3. मिश्रा, एस. डिजिटल पुस्तकालय प्रणाली, ज्ञानदा प्रकाशन।
4. सिंह, बी. आधुनिक पुस्तकालय एवं सूचना विज्ञान, राज पब्लिशिंग हाउस।
5. त्रिपाठी, ए. सूचना संसाधन एवं पुनर्प्राप्ति, वाणी प्रकाशन।
6. यादव, आर. मशीन पठनीय सूचीकरण, दीप एंड दीप।
7. वर्मा, के. पुस्तकालय सूचना प्रणाली, साहित्य भवन।
8. पाण्डेय, एम. डिजिटल युग में पुस्तकालय, लोकभारती।
9. सिंह, एस. सूचना प्रबंधन एवं मेटाडाटा, शुभी पब्लिकेशन।
10. Library of Congress. MARC 21 Format for Bibliographic Data.
11. Dublin Core Metadata Initiative (DCMI). Dublin Core User Guide.
12. Taylor, A. G. Introduction to Cataloging and Classification.
13. Hider, P. Information Resource Description.
14. Caplan, P. Metadata Fundamentals for All Librarians.
15. Baca, M. Introduction to Metadata.
16. Tillett, B. What is FRBR? Library of Congress.

17. Coyle, K. Understanding the Semantic Web.
18. IFLA. Standards for Digital Libraries.
19. Arms, W. Y. Digital Libraries. MIT Press.

इकाई –10 : सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली: उद्देश्य, प्रकार, संचालन एवं डिजाइन (ISAR Systems : Objectives, Types, Operations and Design)

इकाई की रूपरेखा

10.1 प्रस्तावना

10.2 उद्देश्य

10.3 ISAR प्रणाली: अर्थ एवं अवधारणा

10.4 सूचना भंडारण एवं पुनर्प्राप्ति का विकास

10.5 ISAR प्रणालियों के उद्देश्य एवं कार्य

10.6 ISAR प्रणालियों के प्रकार

10.6.1 पारंपरिक ISAR प्रणाली

10.6.2 कंप्यूटरीकृत ISAR प्रणाली

10.6.3 वेब-आधारित ISAR प्रणाली

10.7 ISAR प्रणाली के प्रमुख घटक

10.8 सूचना भंडारण प्रक्रिया (Storage Process)

10.9 सूचना पुनर्प्राप्ति प्रक्रिया (Retrieval Process)

10.10 ISAR प्रणाली का संचालन (Operations)

10.11 ISAR प्रणाली का डिजाइन एवं संरचना

10.12 ISAR प्रणाली में उपयोगकर्ता इंटरफेस

10.13 ISAR प्रणाली का मूल्यांकन

10.14 व्यावहारिक उदाहरण: एक ISAR सिस्टम का कार्यप्रवाह

10.15 सारांश

10.16 शब्दावली

10.17 निबंधात्मक प्रश्न

10.18 लघु प्रश्न एवं MCQ

10.19 अभ्यास प्रश्नों के उत्तर

10.20 संदर्भ ग्रंथ सूची

10.1 प्रस्तावना (Introduction)

आज सूचना का स्वरूप मुद्रित (Printed) से आगे बढ़कर **ई-पुस्तक, ई-जर्नल, डेटाबेस, डिजिटल रिपॉजिटरी, वेब पेज, मल्टीमीडिया** आदि हो गया है। समस्या अब "सूचना की कमी" नहीं, बल्कि **उपयुक्त सूचना को सही समय पर खोज कर उपलब्ध कराना** है। इसी आवश्यकता से **सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली (Information Storage and Retrieval System—ISAR)** का विकास हुआ। ISAR का लक्ष्य है—

- सूचना को **संगठित (organized)** रूप में सुरक्षित रखना
- उपयोगकर्ता की आवश्यकता (information need) के अनुसार **तेज़, सटीक और प्रासंगिक** सूचना निकालना
- खोज को **सरल, उपयोगकर्ता-अनुकूल (user-friendly)** बनाना

सरल विचार:

"ISAR = सही सूचना + सही उपयोगकर्ता + सही समय + सही रूप"

10.2 उद्देश्य (Objectives)

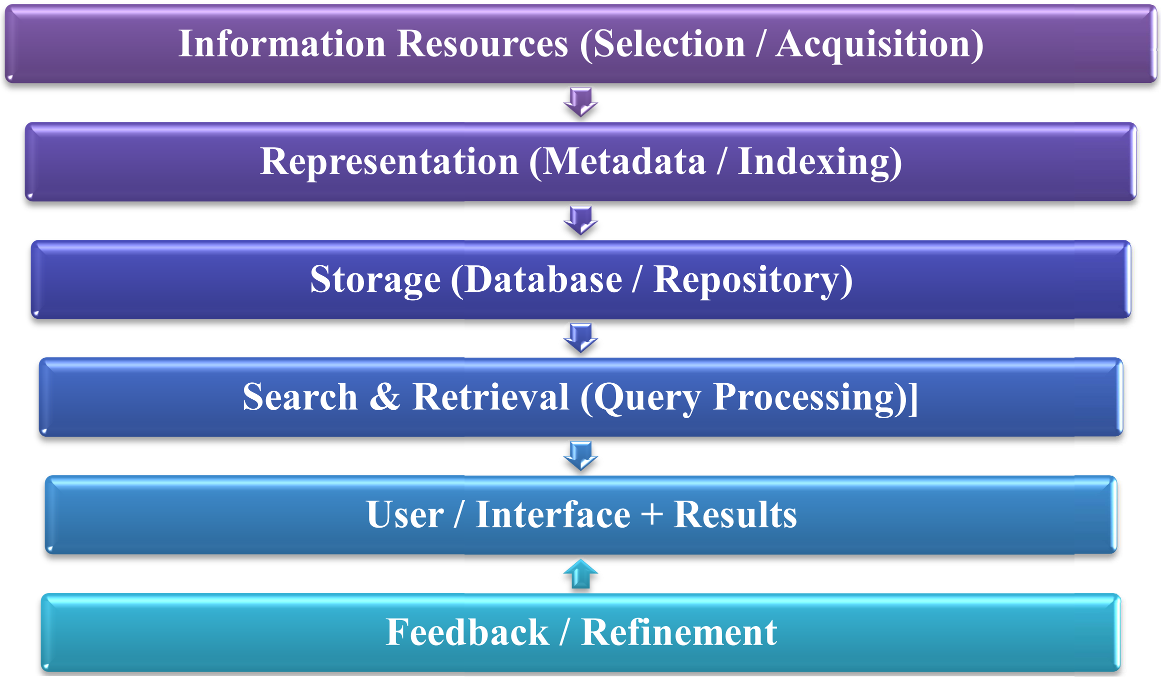
इस इकाई के अध्ययन के बाद मैं/शिक्षार्थी—

1. ISAR प्रणाली की अवधारणा व दायरा समझ सकूँगा/सकूँगी।
2. भंडारण-पुनर्प्राप्ति के विकास चरणों को स्पष्ट कर सकूँगा/सकूँगी।
3. ISAR प्रणालियों के प्रकारों में अंतर बता सकूँगा/सकूँगी।
4. ISAR के घटक, भंडारण व पुनर्प्राप्ति प्रक्रियाएँ समझ सकूँगा/सकूँगी।
5. ISAR के संचालन, डिजाइन और मूल्यांकन के आधार लिख सकूँगा/सकूँगी।
6. व्यावहारिक रूप से एक ISAR workflow/उदाहरण बना सकूँगा/सकूँगी।
7.

10.3 ISAR प्रणाली: अर्थ एवं अवधारणा (Meaning & Concept):

ISAR (Information Storage and Retrieval System) वह प्रणाली है जो सूचना को एक **नियोजित एवं संरचित ढाँचे** में संग्रहित करती है तथा आवश्यकता पड़ने पर **खोज प्रक्रिया (search process)** के माध्यम से उसे **पुनः प्राप्त (retrieve)** कराती है।

ISAR प्रणाली केवल सूचना को संग्रहीत करने तक सीमित नहीं होती, बल्कि यह **सूचना निरूपण (Representation), संगठन (Organization), खोज (Search) तथा पहुँच (Access)** का एक **समन्वित तंत्र** है।



ISAR का मूल मॉडल (Core Model)

10.3 ISAR प्रणाली: अर्थ एवं अवधारणा (Meaning & Concept)

ISAR (Information Storage and Retrieval System) वह प्रणाली है जो सूचना को **एक नियोजित एवं संरचित ढाँचे** में संग्रहित करती है तथा आवश्यकता पड़ने पर **खोज प्रक्रिया (search process)** के माध्यम से उसे **पुनः प्राप्त (retrieve)** कराती है।

ISAR प्रणाली केवल सूचना को संग्रहीत करने तक सीमित नहीं होती, बल्कि यह **सूचना निरूपण (Representation), संगठन (Organization), खोज (Search) तथा पहुँच (Access)** का एक **समन्वित तंत्र** है।

10.4 सूचना भंडारण एवं पुनर्प्राप्ति का विकास (Evolution)

सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली (ISAR) का विकास मुख्यतः निम्नलिखित **चार चरणों** में समझा जा सकता है—

1. **पारंपरिक (Manual) चरण**
   इस चरण में सूचना खोज के लिए **कार्ड-कैटलॉग, शेल्फ-लिस्ट तथा मुद्रित अनुक्रमण एवं सार सेवाओं** का उपयोग किया जाता था।
2. **यांत्रिक / अर्ध-स्वचालित चरण**
   इसमें **माइक्रोफिल्म, पंच-कार्ड** एवं सीमित मशीन सहायता के माध्यम से सूचना खोज की जाती थी।
3. **कंप्यूटरीकृत चरण (Computer-based)**
   इस चरण में **OPAC, बिब्लियोग्राफिक डेटाबेस, की-वर्ड खोज तथा Boolean retrieval** का विकास हुआ, जिससे खोज प्रक्रिया अधिक तेज़ एवं सटीक बनी।
4. **वेब/डिजिटल एवं बुद्धिमान चरण**
   वर्तमान चरण में **सर्च इंजन, डिजिटल लाइब्रेरी, रैंकिंग तकनीक, व्यक्तिगतकरण (personalization) तथा AI-सहायित सूचना पुनर्प्राप्ति** का व्यापक उपयोग हो रहा है।

10.5 ISAR प्रणालियों के उद्देश्य एवं कार्य (Objectives & Functions)

(A) प्रमुख उद्देश्य (Objectives)

- उपयोगकर्ता की सूचना-आवश्यकताओं को समझकर **प्रासंगिक (relevant) सूचना** उपलब्ध कराना
- सूचना खोज में लगने वाले **समय एवं श्रम को कम करना**

- उपलब्ध सूचना संसाधनों का **अधिकतम एवं प्रभावी उपयोग** सुनिश्चित करना
- सूचना संगठन में **मानकीकरण (standards)** एवं **एकरूपता (consistency)** बनाए रखना

(B) मुख्य कार्य (Core Functions)

1. **Acquisition / Collection** – सूचना संसाधनों का चयन एवं संग्रह
2. **Representation** – मेटाडाटा निर्माण, कैटलॉगिंग एवं अनुक्रमण
3. **Organization** – वर्गीकरण, विषय नियंत्रण एवं authority control
4. **Storage** – डेटाबेस/रिपॉजिटरी में सूचना का सुरक्षित भंडारण
5. **Query Processing** – उपयोगकर्ता की query को समझना एवं संशोधित करना
6. **Retrieval & Display** – उपयुक्त परिणाम (records / full-text) प्रस्तुत करना
7. **Feedback & Refinement** – उपयोगकर्ता प्रतिक्रिया के आधार पर खोज में सुधार

10.6 ISAR प्रणालियों के प्रकार (Types of ISAR Systems)

ISAR प्रणालियों को उनके संचालन माध्यम और तकनीकी आधार के अनुसार मुख्यतः पारंपरिक (Manual) और कंप्यूटरीकृत (Computerized) रूप में वर्गीकृत किया जाता है। प्रत्येक प्रकार की अपनी उपयोगिता, विशेषताएँ और सीमाएँ होती हैं।

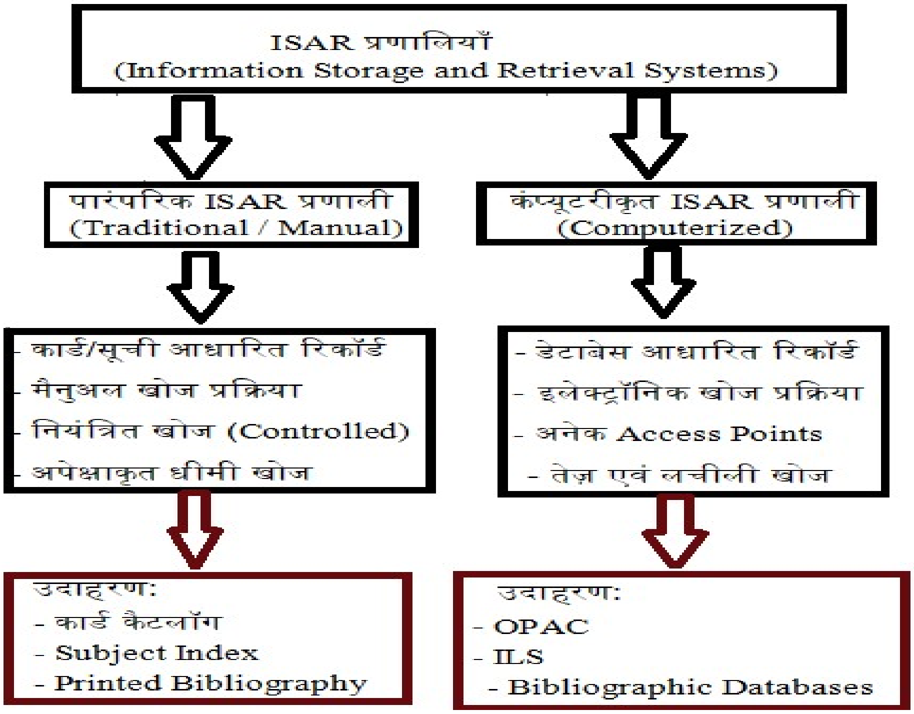


Diagram for 10.6 ISAR प्रणालियाँ

10.6.1 पारंपरिक ISAR प्रणाली (Traditional / Manual Systems)

लक्षण:

- इस प्रकार की प्रणाली में सूचना रिकॉर्ड **मैनुअल रूप** में तैयार एवं संरक्षित किए जाते हैं, जैसे कार्ड, रजिस्टर, शेल्फ-लिस्ट या मुद्रित सूचियाँ।
- खोज प्रक्रिया अपेक्षाकृत **धीमी** होती है, परन्तु यह **नियंत्रित (controlled)** होती है क्योंकि प्रविष्टियाँ निश्चित नियमों और शीर्षकों (controlled headings) के आधार पर बनाई जाती हैं।
- खोज सामान्यतः **लेखक, शीर्षक या विषय** के आधार पर की जाती है तथा एक समय में सीमित access points उपलब्ध होते हैं।

उदाहरण:

- **कार्ड कैटलॉग**
- **Subject Index / Subject headings आधारित सूचियाँ**
- **Printed Bibliography / मुद्रित अनुक्रमण-सार सेवाएँ**

10.6.2 कंप्यूटरीकृत ISAR प्रणाली (Computerized Systems)

लक्षण:

- इस प्रकार की प्रणाली में रिकॉर्ड **डेटाबेस आधारित (database-driven)** होते हैं और इलेक्ट्रॉनिक रूप में संग्रहीत किए जाते हैं।
- उपयोगकर्ता को **अनेक access points** उपलब्ध होते हैं—Author, Title, Subject, Keyword आदि—जिससे खोज अधिक लचीली और प्रभावी होती है।
- खोज प्रक्रिया **तेज़**, अधिक सटीक, और बहु-विध (multi-search options) होती है; जैसे advanced search, filters, sorting आदि।

उदाहरण:

- **OPAC (Online Public Access Catalogue)**
- **Library Management System (ILS)**
- **Bibliographic Databases** (जिनमें रिकॉर्ड और खोज सुविधा दोनों होती हैं)

10.7 ISAR प्रणाली के प्रमुख घटक (Major Components)

ISAR प्रणाली विभिन्न घटकों के समन्वय से कार्य करती है। इन घटकों को मुख्यतः चार श्रेणियों में वर्गीकृत किया जा सकता है—

(A) संसाधन / डेटा घटक

- **सूचना संसाधन**: पुस्तकें, लेख, वेब पृष्ठ, डिजिटल दस्तावेज़ आदि
- **मेटाडाटा एवं सहायक उपकरण**: Metadata, Indexes, Abstracts

(B) तकनीकी घटक

- **Database / Repository** – सूचना का भंडारण स्तर (Storage layer)
- **Indexing Engine / Search Engine** – खोज एवं पुनर्प्राप्ति स्तर (Retrieval layer)

(C) मानव / प्रबंधन घटक

- **मानव संसाधन**: Indexer, Cataloger, System Administrator
- **नीतियाँ**: Indexing policy, Update policy, Access policy

(D) उपयोगकर्ता घटक

- **उपयोगकर्ता एवं इंटरफेस**: User, User Interface
- **सहायता एवं प्रतिक्रिया**: Help/Guidance tools, Feedback mechanism

Users ↔ Interface ↔ Search Engine ↔ Index / Database ↔ Resources
↑
(Admin / Policies / Updates)

ISAR प्रणाली के घटकों का सरल आरेख (Component Diagram)

10.8 सूचना भंडारण प्रक्रिया (Storage Process)

सूचना भंडारण का अर्थ केवल दस्तावेज़ों को **सेव (save)** करना नहीं है, बल्कि उन्हें इस प्रकार **संरचित और व्यवस्थित** करना है कि भविष्य में आवश्यकता पड़ने पर वे **आसानी से खोजी और पुनः प्राप्त (retrievable)** की जा सकें। एक प्रभावी ISAR प्रणाली में भंडारण प्रक्रिया सूचना पुनर्प्राप्ति की गुणवत्ता को प्रत्यक्ष रूप से प्रभावित करती है।

Storage Process के प्रमुख चरण

1. **Selection / Acquisition (चयन एवं अधिग्रहण)**
   इस चरण में पुस्तकें, लेख, डिजिटल दस्तावेज़ अथवा वेब संसाधनों का **उपयुक्त चयन और संग्रह** किया जाता है।

2. **Content Analysis (सामग्री विश्लेषण):** चयनित संसाधनों की विषय-वस्तु का विश्लेषण कर **मुख्य विषय, अवधारणाएँ** एवं **की-वर्ड्स** की पहचान की जाती है।

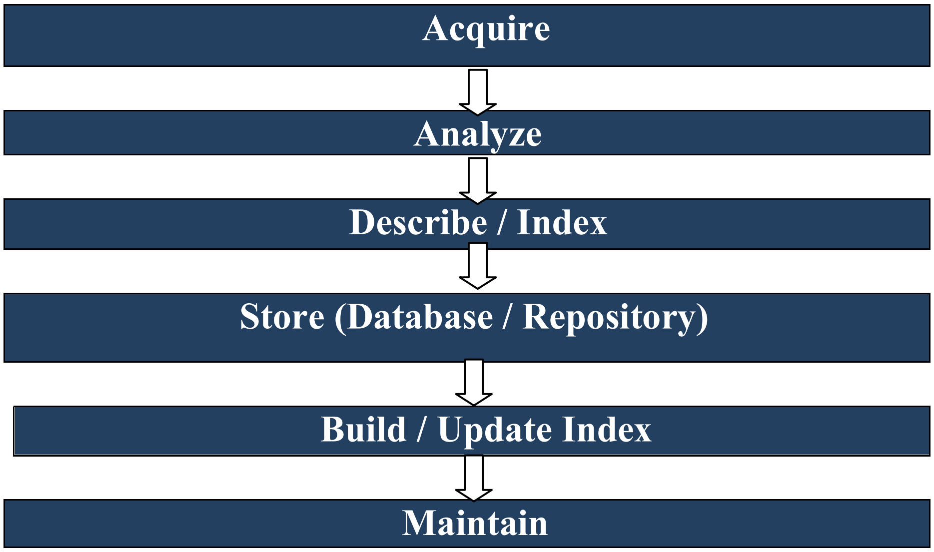
3. **Representation (प्रतिनिधित्व):** संसाधन को **Metadata, Cataloging और Indexing terms** के माध्यम से निरूपित किया जाता है ताकि वह खोज-योग्य बन सके।

4. **Authority / Control (नियंत्रण एवं एकरूपता):** नामों, विषय शीर्षकों और शब्दावली में एकरूपता **(consistency)** बनाए रखने के लिए authority control लागू किया जाता है।

5. **Database Entry (डेटाबेस प्रविष्टि):** तैयार किए गए रिकॉर्ड को **डेटाबेस या डिजिटल रिपॉजिटरी** में संग्रहीत किया जाता है।

6. **Index Build / Update (इंडेक्स निर्माण/अद्यतन):** प्रभावी खोज के लिए **इंडेक्स का निर्माण या अद्यतन** किया जाता है।

7. **Maintenance (रख-रखाव):** इस चरण में रिकॉर्ड का **संशोधन, अद्यतन, विलोपन (deletion)** तथा **संस्करण नियंत्रण (versioning)** किया जाता है।



<u>Fig.: सूचना भंडारण प्रक्रिया का प्रवाह आरेख (Storage Flow Diagram)</u>

10.9 सूचना पुनर्प्राप्ति प्रक्रिया (Retrieval Process)

सूचना पुनर्प्राप्ति प्रक्रिया का मुख्य उद्देश्य उपयोगकर्ता की **सूचना-आवश्यकता (information need)** के अनुरूप **प्रासंगिक (relevant) परिणाम** उपलब्ध कराना है। एक प्रभावी ISAR प्रणाली में पुनर्प्राप्ति प्रक्रिया खोज की सटीकता और उपयोगकर्ता संतुष्टि को सीधे प्रभावित करती है।

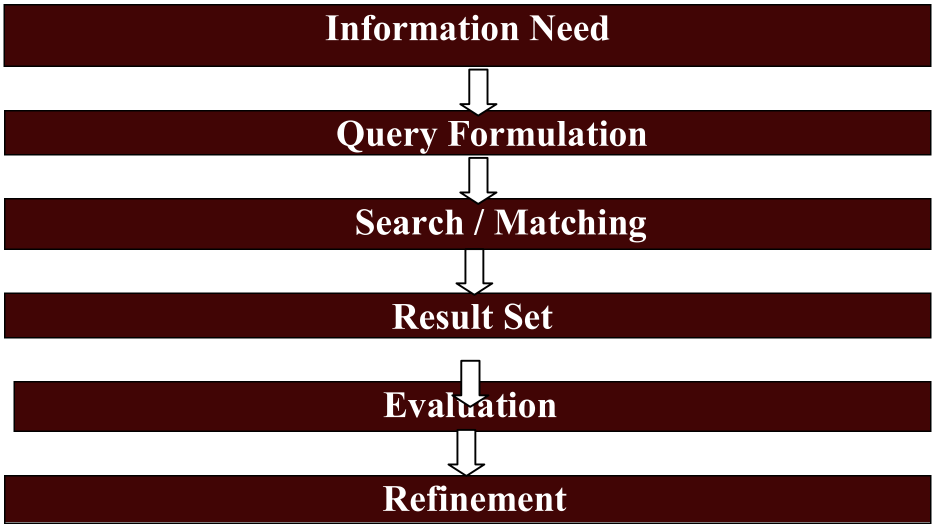


Fig: सूचना पुनर्प्राप्ति प्रक्रिया का प्रवाह आरेख (Retrieval Flow Diagram)

Retrieval Process के प्रमुख चरण

1. **Information Need (सूचना आवश्यकता):** इस चरण में उपयोगकर्ता यह स्पष्ट करता है कि उसे किस प्रकार की सूचना चाहिए।
2. **Query Formulation (खोज प्रश्न निर्माण):** उपयोगकर्ता अपनी आवश्यकता को **keywords या phrases** के रूप में व्यक्त करता है।
3. **Search Strategy (खोज रणनीति):** उपयुक्त खोज विधियों का चयन किया जाता है, जैसे— field search, filters, Boolean operators (AND, OR, NOT) आदि।
4. **Matching (मिलान):**खोज प्रश्न का **इंडेक्स या डेटाबेस** में उपलब्ध रिकॉर्ड से मिलान किया जाता है।
5. **Result Set (परिणाम समूह):** उपयोगकर्ता को **रिकॉर्ड या full-text दस्तावेज़ों** के रूप में परिणाम प्राप्त होते हैं।
6. **Refinement (खोज परिष्करण):** प्राप्त परिणामों के आधार पर खोज को और बेहतर बनाने हेतु filters, synonyms, narrower/broader terms का उपयोग किया जाता है।

उदाहरण (Example)

यदि मुझे **“digital library”** विषय पर लेख खोजने हैं—

- **Query:** "digital library" AND metadata
- **Filters:**
    - Year: 2020–2025
    - Document type: Article
    - Language: English / Hindi
- **Refinement:**
    - "institutional repository" OR DSpace जोड़कर खोज को और सटीक बनाया जाता है।

10.10 ISAR प्रणाली का संचालन (Operations)

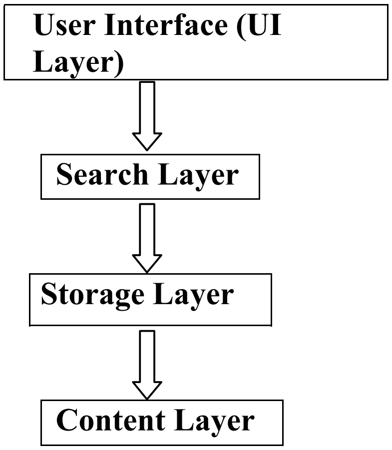
ISAR प्रणाली का संचालन उन दैनिक प्रक्रियाओं का समुच्चय है, जिनके माध्यम से सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली सुचारु रूप से कार्य करती है। ये प्रक्रियाएँ खोज की गुणवत्ता, प्रणाली की कार्यक्षमता तथा उपयोगकर्ता संतुष्टि सुनिश्चित करती हैं।

ISAR संचालन की प्रमुख प्रक्रियाएँ

- **Indexing / Updating:** नए सूचना रिकॉर्ड जोड़ना, पुराने रिकॉर्ड में संशोधन करना तथा इंडेक्स को अद्यतन रखना।
- **Query Processing:** उपयोगकर्ता की खोज-प्रश्न (query) को संसाधित करना, जैसे— stop-words हटाना, stemming करना तथा समानार्थक शब्दों (synonyms) का उपयोग।
- **Searching & Ranking:** query का डेटाबेस/इंडेक्स से मिलान करना तथा प्रासंगिकता के आधार पर परिणामों को क्रमबद्ध (ranking) करना।
- **Access / Delivery:** उपयुक्त रिकॉर्ड या **full-text दस्तावेज़** तक पहुँच उपलब्ध कराना, जैसे link resolver या download सुविधा।
- **Logging / Analytics:** उपयोगकर्ता खोज व्यवहार का विश्लेषण करना, जिससे प्रणाली में सुधार एवं बेहतर सेवाएँ प्रदान की जा सकें।
- **Security & Rights Management:** उपयोगकर्ता अधिकारों का नियंत्रण, लाइसेंस प्रबंधन तथा सुरक्षित पहुँच सुनिश्चित करना।

10.11 ISAR प्रणाली का डिजाइन एवं संरचना (Design & Architecture)

एक प्रभावी ISAR प्रणाली का डिजाइन **उपयोगकर्ता-केंद्रित (user-centered)** तथा **उद्देश्य-केंद्रित (goal-oriented)** होता है। ISAR का डिजाइन इस प्रकार होना चाहिए कि वह उपयोगकर्ता की सूचना-आवश्यकताओं को शीघ्र, सटीक और प्रभावी रूप से पूरा कर सके।



ISAR प्रणाली की संरचना का सरल आरेख (Architecture Diagram)

ISAR डिजाइन के प्रमुख आधार

1. **User Needs Analysis:** उपयोगकर्ताओं की प्रकृति, उनकी सूचना-आवश्यकताओं तथा खोज व्यवहार का विश्लेषण।
2. **Collection / Content Scope:** यह निर्धारण कि प्रणाली में किस प्रकार की सूचना शामिल की जाएगी और किसे बाहर रखा जाएगा।
3. **Metadata एवं Indexing Policy:** नियंत्रित शब्दावली (controlled vocabulary) अथवा की-वर्ड आधारित अनुक्रमण का चयन।
4. **Database Structure:** रिकॉर्ड फील्ड्स, उनके आपसी संबंध तथा पहचानकर्ता (identifiers) का निर्धारण।
5. **Search Features:** Basic एवं advanced search, facets, filters तथा sorting जैसी सुविधाएँ।
6. **Performance:** प्रणाली की गति (speed), क्षमता विस्तार (scalability) एवं स्थिरता।
7. **Interoperability:** मानकों का उपयोग (जैसे MARC, Dublin Core) तथा OAI-PMH के माध्यम से डेटा साझा करने की क्षमता।

8. **Maintenance Plan:** रिकॉर्ड अद्यतन, बैकअप, दीर्घकालिक संरक्षण (digital preservation) की योजना।

10.12 ISAR प्रणाली में उपयोगकर्ता इंटरफेस (User Interface)

उपयोगकर्ता इंटरफेस (UI) ISAR प्रणाली का **"चेहरा"** होता है, जिसके माध्यम से उपयोगकर्ता प्रणाली से संपर्क करता है। एक प्रभावी ISAR प्रणाली में UI को इस प्रकार डिज़ाइन किया जाता है कि वह खोज प्रक्रिया को **सरल, स्पष्ट और उपयोगकर्ता-अनुकूल** बनाए।

एक अच्छे ISAR UI की प्रमुख विशेषताएँ

- **सरलता (Simplicity)** – उपयोग में आसान एवं सहज
- **स्पष्टता (Clear Labels)** – स्पष्ट विकल्प और निर्देश
- **सहायता सुविधा (Help / Tips)** – खोज में मार्गदर्शन
- **फिल्टर एवं फैसेट्स** – परिणामों को सीमित एवं परिष्कृत करने की सुविधा
- **बहुभाषिक समर्थन (Multi-lingual Support)** – जहाँ आवश्यक हो

अच्छे UI का उदाहरण (Library OPAC / Repository)

- Search box के साथ **Advanced Search**
- **Filters: Author / Year / Document Type**
- **Sorting: Relevance / Date**

10.13 ISAR प्रणाली का मूल्यांकन (Evaluation)

ISAR प्रणाली का मूल्यांकन यह निर्धारित करता है कि कोई सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली **कितनी प्रभावी, उपयोगी और उपयोगकर्ता-अनुकूल** है। मूल्यांकन के माध्यम से प्रणाली की गुणवत्ता तथा सुधार की संभावनाओं का आकलन किया जाता है।

ISAR मूल्यांकन के प्रमुख आधार

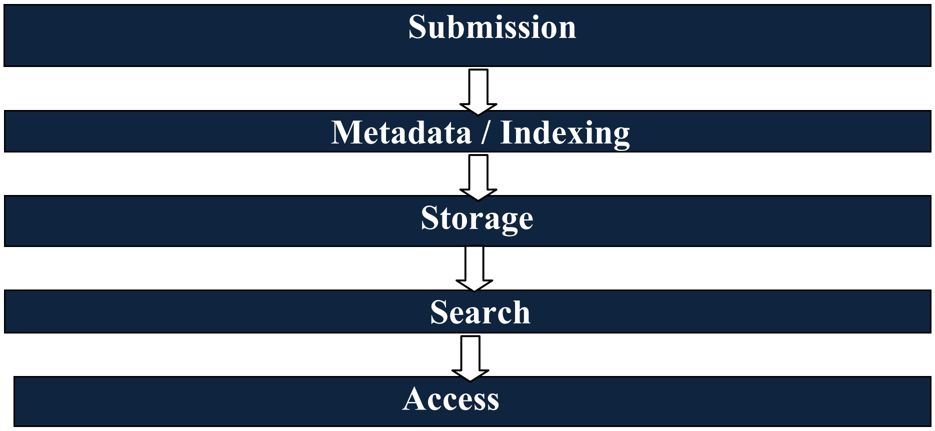
- **Relevance Quality** – प्राप्त परिणाम उपयोगकर्ता की आवश्यकता से कितने प्रासंगिक हैं।
- **Response Time** – खोज परिणाम कितनी शीघ्रता से उपलब्ध होते हैं।
- **Coverage** – प्रणाली में उपलब्ध सूचना संग्रह की उपयुक्तता एवं व्यापकता।

- **User Satisfaction** – उपयोगकर्ता प्रतिक्रिया एवं संतुष्टि स्तर।
- **Usability** – उपयोगकर्ता इंटरफेस की सरलता एवं सहजता।
- **Update Frequency** – सूचना/रिकॉर्ड कितनी नियमितता से अद्यतन किए जाते हैं।

10.14 व्यावहारिक उदाहरण: एक ISAR सिस्टम का कार्यप्रवाह (Workflow)

उदाहरण: Institutional Repository (DSpace-प्रकार)

एक डिजिटल रिपॉजिटरी आधारित ISAR प्रणाली में सूचना भंडारण एवं पुनर्प्राप्ति की प्रक्रिया एक क्रमबद्ध कार्यप्रवाह के माध्यम से सम्पन्न होती है, जिससे सूचना को संगठित रूप में सुरक्षित रखा जा सके और आवश्यकता पड़ने पर उसे आसानी से खोजा जा सके।



ISAR कार्यप्रवाह का सरल आरेख (Workflow Diagram)

1. **Faculty Submission:** – शोधकर्ता अथवा शिक्षक द्वारा शोध-पत्र/डिजिटल दस्तावेज़ रिपॉजिटरी में जमा किया जाता है।
2. **Metadata Creation** :– पुस्तकालयकर्मी संसाधन का **मेटाडाटा** (जैसे शीर्षक, लेखक, की-वर्ड्स, विषय) तैयार करता है, जिससे दस्तावेज़ खोज-योग्य बन सके।
3. **Indexing** :– प्रणाली मेटाडाटा के आधार पर रिकॉर्ड का **इंडेक्स** तैयार करती है।
4. **Searching** :– उपयोगकर्ता उपयुक्त खोज-प्रश्न दर्ज करता है, जैसे—**"open access" AND repository**
5. **Result Display:** – खोज के परिणामस्वरूप उपयोगकर्ता को रिकॉर्ड के साथ **PDF लिंक** दिखाई देता है।
6. **Access / Use:** – उपयोगकर्ता दस्तावेज़ को **preview या download** कर सकता है।

7. **Logging / Feedback:** –प्रणाली उपयोग के आँकड़े (usage statistics) और प्रतिक्रिया दर्ज करती है, जिससे सेवा में सुधार किया जा सके।

10.15 सारांश (Summary)

इस इकाई में **सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली (ISAR)** की अवधारणा, विकास, प्रकार, प्रमुख घटक, भंडारण एवं पुनर्प्राप्ति प्रक्रियाएँ, संचालन, डिजाइन तथा मूल्यांकन के मूलभूत पक्षों का अध्ययन किया गया। यह स्पष्ट हुआ कि ISAR प्रणाली केवल सूचना के संग्रह तक सीमित नहीं है, बल्कि वह सूचना के **संगठन, खोज और प्रभावी उपयोग** की एक समन्वित व्यवस्था है।

निष्कर्षतः ISAR प्रणाली आधुनिक पुस्तकालयों एवं डिजिटल सूचना सेवाओं की **"कार्यात्मक रीढ़"** के रूप में कार्य करती है, जो सूचना को **संगठित, खोज-योग्य, सुलभ और उपयोगकर्ता-उन्मुख** बनाकर ज्ञान के प्रभावी प्रसार में महत्वपूर्ण भूमिका निभाती है।

10.16 शब्दावली (Glossary)

- **ISAR (Information Storage and Retrieval System)** सूचना को संगठित रूप में संग्रहित कर आवश्यकता पड़ने पर उसे खोजने एवं पुनः प्राप्त करने वाली प्रणाली।
- **Indexing (अनुक्रमण):** दस्तावेज़ की विषय-वस्तु को प्रतिनिधि शब्दों/टर्म्स में व्यक्त करने की प्रक्रिया, जिससे खोज संभव हो सके।
- **Query (खोज-प्रश्न):** उपयोगकर्ता द्वारा सूचना प्राप्त करने हेतु दर्ज किया गया खोज कथन या शब्द-समूह।
- **Repository (डिजिटल रिपॉजिटरी):** डिजिटल दस्तावेज़ों को मेटाडाटा सहित संग्रहीत, संरक्षित और उपलब्ध कराने वाला मंच।
- **Interoperability (अंतर-संचालनीयता):** विभिन्न प्रणालियों के बीच डेटा और सेवाओं के आदान-प्रदान की क्षमता।
- **User Interface (UI):** वह माध्यम जिसके द्वारा उपयोगकर्ता प्रणाली से संवाद करता है।
- **Metadata (मेटाडाटा):** "डेटा के बारे में डेटा", जो किसी संसाधन के विवरण, पहचान और प्रबंधन में सहायक होता है।
- **Retrieval (पुनर्प्राप्ति):** संग्रहीत सूचना को खोज कर उपयोगकर्ता तक पहुँचाने की प्रक्रिया।

- **Authority Control:** नामों और विषय शीर्षकों में एकरूपता बनाए रखने की व्यवस्था।

10.17 निबंधात्मक प्रश्न – उत्तर (Model Answers)

प्रश्न 1. ISAR प्रणाली की अवधारणा समझाते हुए इसके उद्देश्य व कार्यों का विवेचन कीजिए।

उत्तर:
ISAR (Information Storage and Retrieval System) वह प्रणाली है जो सूचना को संगठित रूप में संग्रहित करती है तथा उपयोगकर्ता की आवश्यकता के अनुसार उसे पुनः प्राप्त करती है। इसका मुख्य उद्देश्य सही सूचना को सही समय पर सही उपयोगकर्ता तक पहुँचाना है। ISAR के प्रमुख कार्यों में सूचना का संग्रह, मेटाडाटा निर्माण, अनुक्रमण, भंडारण, खोज, परिणाम प्रस्तुति तथा उपयोगकर्ता प्रतिक्रिया के आधार पर सुधार शामिल हैं।

प्रश्न 2. पारंपरिक, कंप्यूटरीकृत और वेब-आधारित ISAR प्रणालियों की तुलना उदाहरण सहित कीजिए।

उत्तर:
पारंपरिक ISAR प्रणालियाँ मैनुअल साधनों पर आधारित होती हैं, जैसे कार्ड कैटलॉग। कंप्यूटरीकृत ISAR में डेटाबेस और OPAC का उपयोग होता है, जिससे खोज तेज़ होती है। वेब-आधारित ISAR प्रणालियाँ इंटरनेट पर आधारित होती हैं और full-text खोज, दूरस्थ पहुँच तथा लिंक्स की सुविधा देती हैं। उदाहरणस्वरूप—कार्ड कैटलॉग, OPAC और डिजिटल रिपॉजिटरी।

प्रश्न 3. ISAR सिस्टम के प्रमुख घटकों का चित्र सहित वर्णन कीजिए।

उत्तर:
ISAR प्रणाली के प्रमुख घटक हैं—डेटा/संसाधन घटक, तकनीकी घटक, मानव/प्रबंधन घटक और उपयोगकर्ता घटक। ये सभी घटक मिलकर सूचना के भंडारण, खोज और पुनर्प्राप्ति को संभव बनाते हैं।
(परीक्षा में सरल flow diagram बनाया जा सकता है: User → Interface → Search Engine → Database)

प्रश्न 4. ISAR प्रणाली के डिजाइन में “User Needs” की भूमिका स्पष्ट कीजिए।

उत्तर:
ISAR प्रणाली का डिजाइन उपयोगकर्ता की सूचना-आवश्यकताओं पर आधारित होता है।

उपयोगकर्ता कौन है, क्या खोजता है और कैसे खोजता है—इन सभी बातों को ध्यान में रखकर UI, search features और metadata नीति तय की जाती है। इससे प्रणाली अधिक प्रभावी और उपयोगकर्ता-अनुकूल बनती है।

प्रश्न 5. ISAR प्रणाली के मूल्यांकन के प्रमुख आधार लिखिए।

उत्तर:
ISAR प्रणाली का मूल्यांकन relevance quality, response time, coverage, user satisfaction, usability और update frequency जैसे आधारों पर किया जाता है। मूल्यांकन से प्रणाली की प्रभावशीलता और सुधार की दिशा निर्धारित होती है।

10.18 लघु प्रश्न – उत्तर (Short Answers)

1. **ISAR प्रणाली से आप क्या समझते हैं?**
   सूचना को संगठित कर खोज एवं पुनर्प्राप्ति की सुविधा देने वाली प्रणाली।
2. **Storage process के मुख्य चरण लिखिए।**
   Selection, Content analysis, Representation, Database entry, Indexing, Maintenance।
3. **Retrieval process में query refinement क्यों आवश्यक है?**
   प्रासंगिक और सटीक परिणाम प्राप्त करने हेतु।
4. **Web-based ISAR की दो विशेषताएँ लिखिए।**
   Full-text search और दूरस्थ पहुँच।
5. **User interface को user-friendly बनाने के दो उपाय लिखिए।**
   सरल डिज़ाइन और स्पष्ट labels।
6. **ISAR में interoperability का क्या अर्थ है?**
   विभिन्न प्रणालियों के बीच डेटा साझा करने की क्षमता।
7. **OPAC और search engine में एक मुख्य अंतर लिखिए।**
   OPAC नियंत्रित पुस्तकालय संग्रह खोजता है, जबकि search engine वेब सामग्री।
8. **ISAR operations में indexing update का महत्व बताइए।**
   अद्यतन इंडेक्स से खोज परिणाम अधिक सटीक होते हैं।

10.18.1 MCQ

1. ISAR का मुख्य उद्देश्य है—
A) केवल दस्तावेज़ संग्रह करना

B) सही सूचना को सही समय पर उपलब्ध कराना ☑
C) केवल वेबसाइट बनाना
D) केवल पुस्तकें छापना

2. पारंपरिक ISAR का उदाहरण है—
A) Google
B) DSpace
C) Card Catalogue ☑
D) Web crawler

3. कंप्यूटरीकृत ISAR का सामान्य उदाहरण—
A) OPAC ☑
B) मैनुअल शेल्फ लिस्ट
C) मुद्रित बिब्लियोग्राफी
D) हस्तलिखित सूची

4. Web-based ISAR की विशेषता—
A) Remote access ☑
B) केवल कागज़ पर खोज
C) केवल एक access point
D) खोज के बिना परिणाम

5. Storage process में "Representation" का अर्थ—
A) दस्तावेज़ फाड़ना
B) metadata/index terms देना ☑
C) केवल छपाई
D) केवल binding

6. Retrieval में पहला चरण—
A) Result display
B) Information need ☑
C) Index update
D) Catalog printing

7. UI का मुख्य उद्देश्य—
A) उपयोगकर्ता को भ्रमित करना
B) उपयोगकर्ता-सिस्टम संवाद आसान बनाना ☑
C) केवल सर्वर चलाना
D) केवल बैकअप बनाना

8. ISAR design में सबसे पहले—
A) रंग चुनना
B) User needs analysis ☑
C) पोस्टर बनाना
D) केवल डेटा हटाना

9. Interoperability का अर्थ—
A) सिस्टम बंद करना
B) प्रणालियों के बीच संगतता ☑
C) केवल एक कंप्यूटर
D) केवल ऑफलाइन खोज

10. Index का मुख्य कार्य—
A) खोज को तेज़ बनाना ☑
B) डेटा मिटाना
C) केवल प्रिंट करना
D) केवल स्कैन करना

MCQ answer key (1–10): **B, C, A, A, B, B, B, B, B, A**

10.19 संदर्भ ग्रंथ सूची (References)

1. Lancaster, F. W. Information Retrieval Systems: Characteristics, Testing and Evaluation.
2. Chowdhury, G. G. Introduction to Modern Information Retrieval.
3. Baeza-Yates, R., & Ribeiro-Neto, B. Modern Information Retrieval.
4. Salton, G. Automatic Text Processing / The SMART Retrieval System.
5. Manning, C. D., Raghavan, P., & Schütze, H. Introduction to Information Retrieval.
6. Van Rijsbergen, C. J. Information Retrieval.
7. Rowley, J. Information Retrieval and the Internet.
8. Belkin, N. J. (Selected works on IR & information need).
9. Harter, S. P. (Works on IR evaluation and searching).
10. Cleveland, D. B., & Cleveland, A. D. Introduction to Indexing and Abstracting.
11. रँगनाथन, एस. आर. — पुस्तकालय संगठन/सूचना संगठन पर चयनित लेख/ग्रंथ (भारतीय संदर्भ)।
12. कृष्ण कुमार — पुस्तकालय एवं सूचना विज्ञान (भारतीय LIS लेखन/संग्रह).

13. भारतीय राष्ट्रीय/विश्वविद्यालय स्तर की **डिजिटल लाइब्रेरी/रिपॉजिटरी गाइडलाइंस** (INFLIBNET/NDLI आदि) – उपयोगकर्ता व खोज संदर्भ।
14. **INFLIBNET** प्रशिक्षण/दस्तावेज़: OPAC, databases, resource sharing (सहायक संदर्भ)।
15. **DELNET/Union Catalogue** से संबंधित परिचयात्मक सामग्री (resource discovery संदर्भ)।
16. "सूचना पुनर्प्राप्ति" विषय पर **हिंदी माध्यम की MLIS नोट्स/पाठ्य सामग्री** (विश्वविद्यालय प्रकाशन)।
17. **NPTEL/SWAYAM (LIS/IR related lectures)**—संकल्पनात्मक समर्थन हेतु (यदि आपके पाठ्यक्रम में स्वीकार्य हो)।

इकाई–11 : ISAR प्रणालियों की अनुकूलन क्षमता (Adaptability of ISAR Systems)

इकाई की रूपरेखा

11.1 प्रस्तावना

11.2 उद्देश्य

11.3 ISAR प्रणाली में अनुकूलन (Adaptability): अर्थ

11.4 उपयोगकर्ता आवश्यकताओं एवं व्यवहार का प्रभाव

11.5 सूचना परिवेश एवं तकनीकी परिवर्तन

11.6 ISAR प्रणालियों में लचीलापन (Flexibility)

11.7 स्केलेबिलिटी (Scalability) एवं इंटरऑपरेबिलिटी

11.8 बहुभाषिक एवं बहु-विषयक अनुकूलन

11.9 डिजिटल रिपॉजिटरी एवं वेब वातावरण में अनुकूलन

11.10 उपयोगकर्ता-केंद्रित ISAR प्रणालियाँ

11.11 अनुकूलन से संबंधित चुनौतियाँ

11.12 व्यावहारिक उदाहरण: Adaptive ISAR system

11.13 आधुनिक प्रवृत्तियाँ एवं भविष्य की दिशा

11.14 सारांश

11.15 शब्दावली

11.16 निबंधात्मक प्रश्न

11.17 लघु प्रश्न एवं MCQ

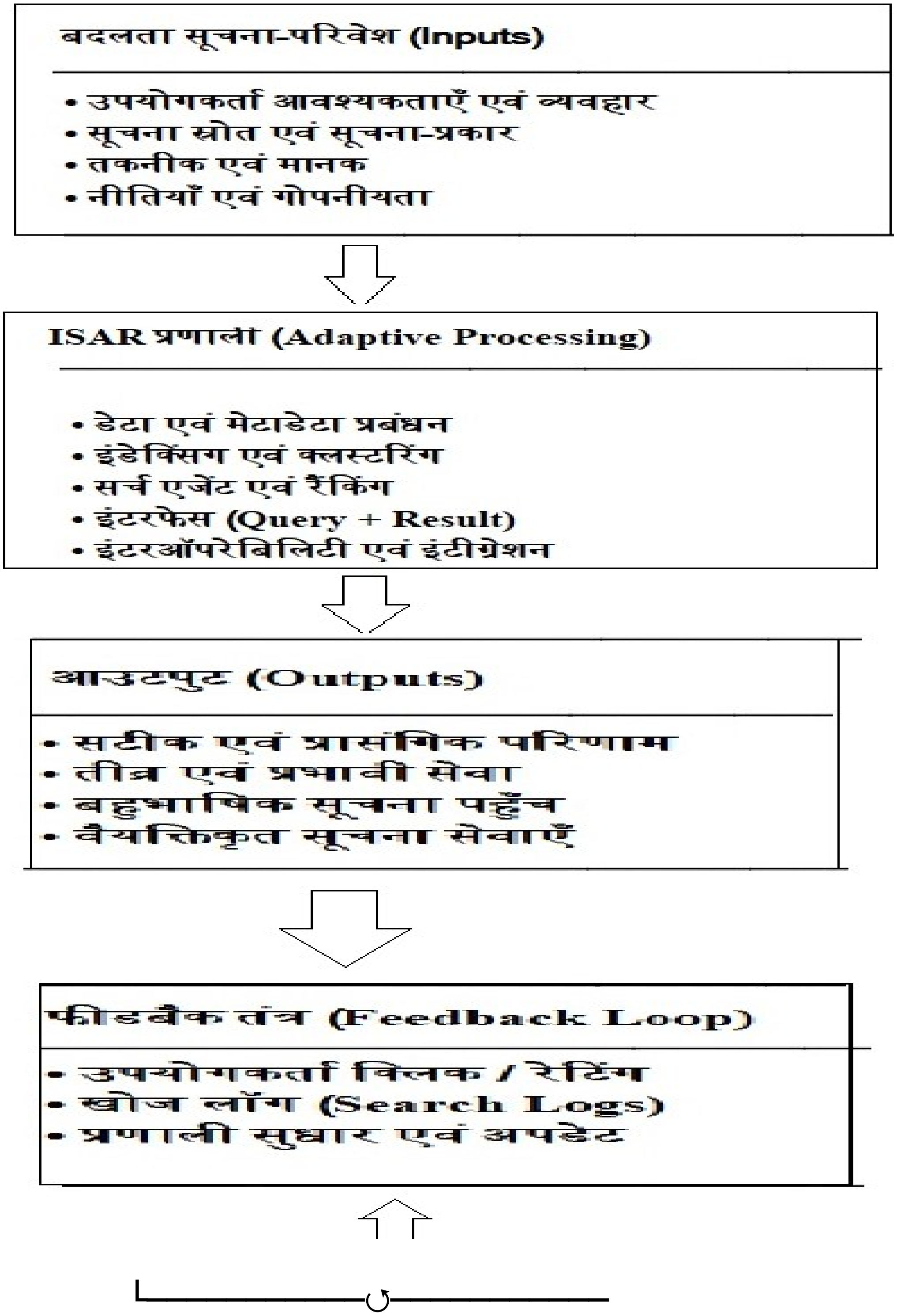
11.18 अभ्यास प्रश्नों के उत्तर

11.19 संदर्भ ग्रंथ सूची

11.1 प्रस्तावना (Introduction) :

आज का सूचना-परिवेश स्थिर (static) नहीं रह गया है, बल्कि यह निरंतर परिवर्तनशील (dynamic) हो चुका है। सूचना अब केवल मुद्रित पुस्तकों या पत्रिकाओं तक सीमित न होकर **डेटासेट, ई-जर्नल (OA), मल्टीमीडिया संसाधन, वेब-सामग्री और डिजिटल रिपॉजिटरी** के रूप में उपलब्ध है। इसी के साथ Cloud computing, AI/ML, semantic web जैसी नवीन तकनीकों तथा **मोबाइल-आधारित खोज, voice search और तत्काल उत्तर (instant information)** जैसी उपयोगकर्ता अपेक्षाओं ने सूचना भंडारण एवं पुनर्प्राप्ति प्रणालियों (ISAR) के स्वरूप को गहराई से प्रभावित किया है।

ऐसे परिवेश में आधुनिक ISAR प्रणालियों का सबसे महत्वपूर्ण गुण बन गया है—**अनुकूलन क्षमता (Adaptability)**। अनुकूलन क्षमता से आशय ISAR प्रणाली की उस योग्यता से है, जिसके द्वारा वह **बदलती उपयोगकर्ता आवश्यकताओं, सूचना-प्रकारों, तकनीकी मानकों और खोज-व्यवहार** के अनुरूप अपने **डेटा-प्रबंधन, इंटरफेस, खोज-रणनीतियों और सेवाओं** को समायोजित कर सके। संक्षेप में, जो ISAR प्रणाली बढ़ते डेटा, नई भाषाओं और नई तकनीकों के साथ भी **तेज़, सटीक और उपयोगकर्ता-अनुकूल** बनी रहती है, वही वास्तव में एक Adaptive ISAR System कहलाती है।



चित्र 11.1 : Adaptive ISAR Framework

यह चित्र दर्शाता है कि ISAR प्रणाली किस प्रकार बदलते सूचना-परिवेश से इनपुट लेकर adaptive processing के माध्यम से उपयुक्त आउटपुट प्रदान करती है तथा उपयोगकर्ता फीडबैक के आधार पर स्वयं को निरंतर अनुकूलित (adapt) करती रहती है।

11.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् आप सक्षम होंगे कि आप:

1. ISAR प्रणालियों में **अनुकूलन क्षमता (Adaptability)** के अर्थ, प्रकृति तथा प्रमुख आयामों को सैद्धान्तिक रूप में समझ सकें।
2. **उपयोगकर्ता आवश्यकताओं (User Needs) एवं उपयोगकर्ता व्यवहार (User Behaviour)** में होने वाले परिवर्तनों का ISAR प्रणालियों पर पड़ने वाले प्रभाव का विश्लेषण कर सकें।
3. **Flexibility, Scalability तथा Interoperability** जैसी अवधारणाओं को ISAR की अनुकूलन क्षमता से जोड़कर उनकी भूमिका स्पष्ट कर सकें।
4. **बहुभाषिक (Multilingual) एवं बहु-विषयक (Multidisciplinary)** परिवेश में ISAR प्रणालियों के अनुकूलन के विभिन्न तरीकों को समझ सकें।
5. **डिजिटल रिपॉजिटरी एवं वेब-आधारित सूचना परिवेश** में अनुकूलनशील ISAR प्रणालियों के व्यावहारिक अनुप्रयोगों एवं मॉडलों की व्याख्या कर सकें।
6. ISAR प्रणालियों में अनुकूलन से संबंधित प्रमुख **चुनौतियों** (जैसे लागत, गोपनीयता, मानकीकरण एवं प्रशिक्षण) तथा **भविष्य की प्रवृत्तियों** (जैसे AI-आधारित खोज एवं semantic search) का समालोचनात्मक अध्ययन कर सकें।

11.3 ISAR प्रणाली में अनुकूलन (Adaptability): अर्थ और अवधारणा (Meaning and Concept of Adaptability in ISAR Systems):

ISAR प्रणाली में **अनुकूलन क्षमता (Adaptability)** से आशय उस योग्यता से है, जिसके माध्यम से कोई सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली बदलते **तकनीकी परिवेश, उपयोगकर्ता आवश्यकताओं, सूचना-प्रकारों तथा खोज-व्यवहार** के अनुरूप स्वयं को समायोजित कर सके। Adaptability को समझने के लिए इसे सामान्यतः तीन प्रमुख स्तरों (layers) में वर्गीकृत किया जाता है, जो मिलकर ISAR प्रणाली को प्रभावी, लचीला और भविष्य-उन्मुख बनाते हैं।

(A) तकनीकी अनुकूलन (Technical Adaptability)

तकनीकी अनुकूलन ISAR प्रणाली की उस क्षमता को दर्शाता है, जिसके द्वारा वह नवीन तकनीकों और प्लेटफॉर्म के साथ स्वयं को संगत (compatible) बना सके। इसमें शामिल हैं—

- **नए Operating System, Server अथवा Cloud वातावरण** को अपनाना, जिससे प्रणाली अधिक scalable और reliable बन सके।
- **नई indexing तकनीकों या search engines** (जैसे तेज़ indexing, बेहतर ranking algorithms) का उपयोग, जिससे सूचना पुनर्प्राप्ति की गति और सटीकता बढ़े।
- **API integration, automation तथा security upgrades**, ताकि ISAR प्रणाली अन्य प्रणालियों से जुड़ सके, प्रक्रियाएँ स्वचालित हों और डेटा की सुरक्षा सुनिश्चित की जा सके।

यह स्तर ISAR की आधारभूत संरचना (infrastructure) को मजबूत बनाता है और प्रणाली को तकनीकी रूप से अप्रचलित होने से बचाता है।

(B) कार्यात्मक अनुकूलन (Functional Adaptability)

कार्यात्मक अनुकूलन ISAR प्रणाली की सेवाओं और कार्यों (functions) में होने वाले लचीलेपन से संबंधित है। इसके अंतर्गत—

- **Simple search के साथ-साथ Advanced search सुविधाएँ** (Boolean operators, filters, facets आदि) उपलब्ध कराना, ताकि विभिन्न प्रकार के उपयोगकर्ताओं की खोज-आवश्यकताएँ पूरी हो सकें।
- **Alerts, recommendation services तथा personalization**, जिससे उपयोगकर्ता को उसकी रुचि के अनुसार सूचना स्वतः उपलब्ध हो सके।
- **नई सामग्री-प्रकारों** (text, image, audio, video आदि) की प्रभावी retrieval, जिससे ISAR प्रणाली बहु-माध्यमीय (multimedia-oriented) बन सके।

कार्यात्मक अनुकूलन ISAR प्रणाली को उपयोगकर्ता-उपयोगी और सेवा-उन्मुख बनाता है।

(C) उपयोगकर्ता-केंद्रित अनुकूलन (User-Centred Adaptability)

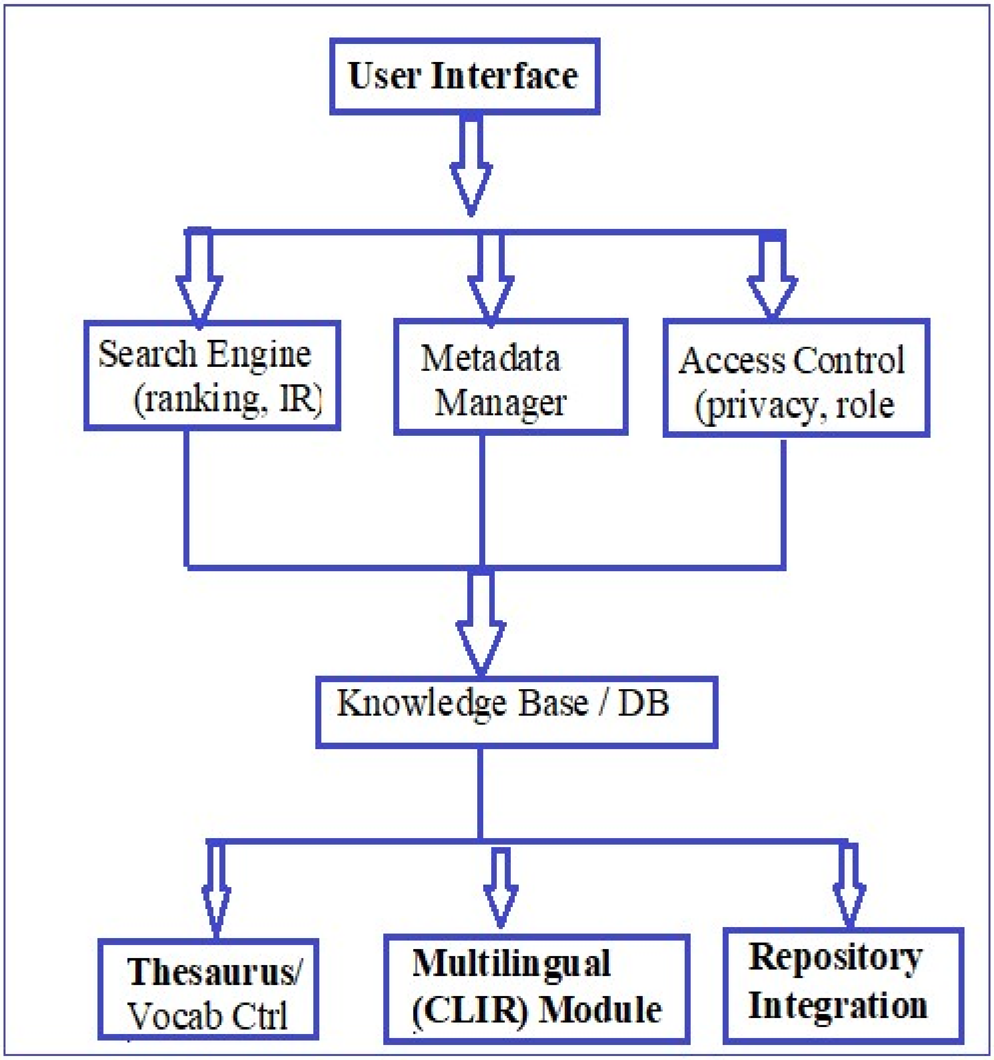
उपयोगकर्ता-केंद्रित अनुकूलन ISAR प्रणाली के मानवीय पक्ष (human dimension) पर केंद्रित होता है। इसमें—

- **UI/UX को सरल, सहज और समझने योग्य बनाना**, ताकि तकनीकी ज्ञान कम होने पर भी उपयोगकर्ता आसानी से प्रणाली का उपयोग कर सके।

- **उपयोगकर्ता को query refine करने में सहायता** देना, जैसे सुझाव, controlled vocabulary या auto-completion, जिससे खोज अधिक सटीक हो।
- **परिणामों की display को "समझने योग्य" और "actionable" बनाना**, जैसे relevance ranking, clear metadata, download/link options आदि।

यह स्तर ISAR प्रणाली की स्वीकृति (acceptability) और प्रभावशीलता को बढ़ाता है।

अधिकांश आधुनिक ISAR प्रणालियाँ **modular basis** पर डिज़ाइन की जाती हैं, ताकि विभिन्न components या subsystems को आवश्यकता अनुसार बदला, जोड़ा या अपडेट किया जा सके। यही modular संरचना ISAR प्रणाली की **अनुकूलन क्षमता को प्रत्यक्ष रूप से समर्थन** प्रदान करती है और उसे बदलते सूचना परिवेश में टिकाऊ (sustainable) बनाती है।

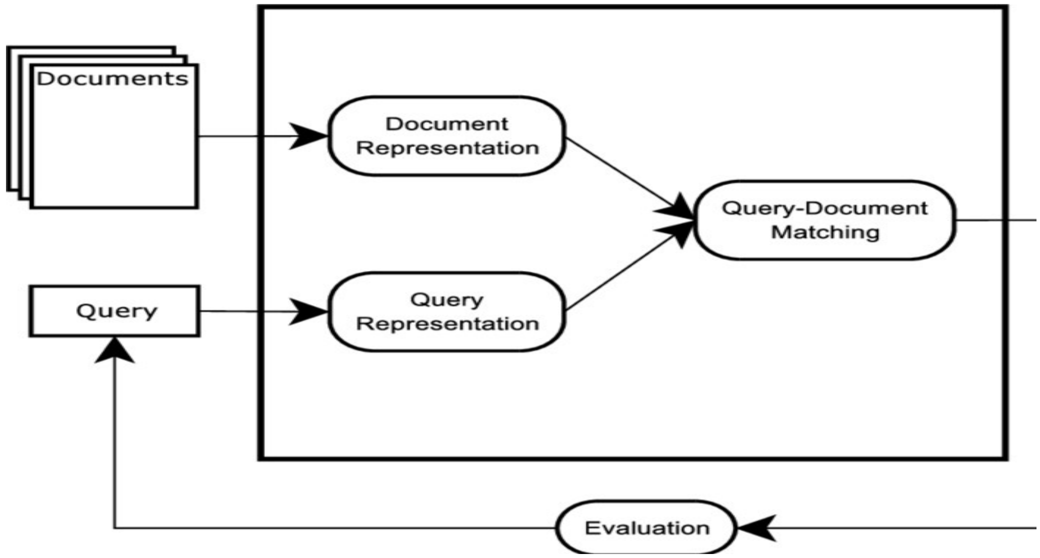


चित्र 11.1 : Adaptive ISAR Framework

Table 11.2 : ISAR में अनुकूलन के प्रमुख स्तरों का तुलनात्मक विवरण

(Comparative View of Adaptability in ISAR Systems)

अनुकूलन का स्तर	मुख्य फोकस	प्रमुख विशेषताएँ	ISAR पर प्रभाव
तकनीकी अनुकूलन **(Technical Adaptability)**	तकनीकी अवसंरचना (Infrastructure)	नया OS/Server/Cloud अपनाना, indexing engine का उन्नयन, API integration, automation, security upgrades	प्रणाली को **scalable, secure** और **future-ready** बनाता है
कार्यात्मक अनुकूलन **(Functional Adaptability)**	सेवाएँ एवं कार्य (Functions & Services)	Simple व Advanced search, Boolean operators, alerts, recommendation, multimedia retrieval	**सूचना पुनर्प्राप्ति को अधिक सटीक, तेज़ और उपयोगी बनाता है**
उपयोगकर्ता-केंद्रित अनुकूलन **(User-Centred Adaptability)**	उपयोगकर्ता अनुभव (User Experience)	सरल UI/UX, query refinement support, स्पष्ट व actionable result display	प्रणाली की **user acceptance, satisfaction** और **usability** बढ़ाता है



11.3 के लिए Key Points

- ISAR प्रणाली में **Adaptability** का अर्थ केवल तकनीकी उन्नयन नहीं, बल्कि **तकनीक, कार्यक्षमता और उपयोगकर्ता अनुभव**—तीनों स्तरों पर समायोजन क्षमता है।
- **Technical adaptability** ISAR को नई तकनीकों और प्लेटफॉर्म के साथ संगत बनाए रखती है।
- **Functional adaptability** विभिन्न प्रकार की खोज आवश्यकताओं और सूचना-प्रकारों को पूरा करने में सहायक होती है।
- **User-centred adaptability** ISAR प्रणाली को सरल, सहज और उपयोगकर्ता-अनुकूल बनाती है।
- **Modular design** ISAR प्रणालियों में अनुकूलन क्षमता का आधार है, क्योंकि इससे components/subsystems को आसानी से बदला या अपडेट किया जा सकता है।
- प्रभावी Adaptability के बिना ISAR प्रणाली **धीरे-धीरे अप्रासंगिक (obsolete)** हो सकती है।
- इस प्रकार, Adaptability ISAR प्रणाली की **दीर्घकालिक उपयोगिता (sustainability)** और **प्रभावशीलता (effectiveness)** का प्रमुख सूचक है।

11.4 उपयोगकर्ता आवश्यकताओं एवं व्यवहार का प्रभाव (Impact of User Needs and User Behaviour on ISAR Systems)

आधुनिक सूचना-परिवेश में ISAR प्रणालियों का स्वरूप मुख्यतः **उपयोगकर्ता आवश्यकताओं (User Needs)** और **उपयोगकर्ता व्यवहार (User Behaviour)** से प्रभावित होता है। आज का उपयोगकर्ता केवल सूचना की उपलब्धता से संतुष्ट नहीं होता, बल्कि वह **तेज़, सटीक और संदर्भ-आधारित (context-aware)** सूचना की अपेक्षा करता है। डिजिटल एवं वेब-आधारित वातावरण में उपयोगकर्ता प्रायः निम्नलिखित विशेषताओं की माँग करता है—

- **Quick answers** अर्थात न्यूनतम समय में प्रासंगिक परिणाम
- **मोबाइल-अनुकूल (mobile-friendly) इंटरफेस**, ताकि किसी भी स्थान से सूचना प्राप्त की जा सके
- **Context-aware परिणाम**, जो उपयोगकर्ता की रुचि, खोज-इतिहास या उद्देश्य के अनुरूप हों
- **Saved searches, alerts एवं notifications**, जिससे नई सूचना स्वतः प्राप्त होती रहे

इन अपेक्षाओं के कारण ISAR प्रणालियों को पारंपरिक “one-size-fits-all” दृष्टिकोण से आगे बढ़कर **उपयोगकर्ता-केंद्रित (user-centred)** दृष्टिकोण अपनाना पड़ता है। इसके लिए आधुनिक ISAR प्रणालियाँ **User Profiles**, **Search Logs** तथा **Feedback Loops** का उपयोग करती हैं, जिससे उपयोगकर्ता के खोज-व्यवहार का विश्लेषण कर सेवाओं को बेहतर बनाया जा सके।

उदाहरण (Illustrative Example)

- **शोधार्थी (Researcher)** के लिए ISAR प्रणाली में *advanced filters, Boolean search, subject-based refinement* जैसी सुविधाएँ अधिक उपयोगी होती हैं।
- वहीं **सामान्य नागरिक (General User)** के लिए *simple search box, guided search एवं स्पष्ट परिणाम-प्रदर्शन* अधिक प्रभावी होता है।

इस प्रकार, उपयोगकर्ता आवश्यकताओं एवं व्यवहार की समझ ISAR प्रणालियों को **अधिक प्रभावी, अनुकूलनशील और उपयोगकर्ता-संतोषकारी** बनाने में महत्वपूर्ण भूमिका निभाती है।

11.5 सूचना परिवेश एवं तकनीकी परिवर्तन (Information Environment and Technological Changes)

सूचना परिवेश में होने वाले तकनीकी परिवर्तन ISAR प्रणालियों की संरचना, कार्यप्रणाली और सेवाओं को निरंतर प्रभावित करते हैं। जैसे-जैसे सूचना के संग्रहण, प्रसंस्करण और पुनर्प्राप्ति की तकनीकें विकसित होती हैं, ISAR प्रणालियों को भी अपने ढाँचे और पहुँच-तंत्र में आवश्यक परिवर्तन करना पड़ता है। तकनीकी परिवर्तनों के अनुरूप स्वयं को अनुकूलित न कर पाने वाली ISAR प्रणालियाँ शीघ्र ही अप्रासंगिक (obsolete) हो जाती हैं।

तकनीकी विकास के क्रम में ISAR प्रणालियों का विकास सामान्यतः इस प्रकार देखा जा सकता है—
Offline → CD-ROM → Online → Web-based → Mobile / AI-based Systems
यह क्रम दर्शाता है कि ISAR प्रणालियाँ सीमित और स्थान-आधारित उपयोग से आगे बढ़कर आज **anytime–anywhere access** प्रदान करने लगी हैं।

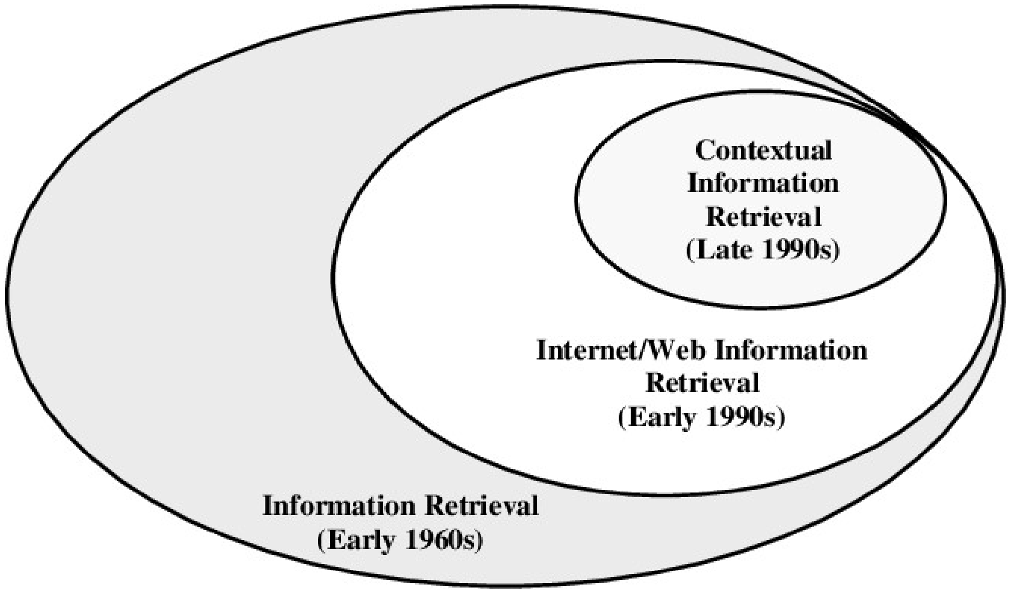
आधुनिक वेब-आधारित सूचना परिवेश में उपयोगकर्ता वेब ब्राउज़र के माध्यम से ISAR प्रणाली का उपयोग करता है, जहाँ ऑपरेटिंग सिस्टम या हार्डवेयर का अंतर कम महत्त्वपूर्ण रह जाता है। यह वेब तकनीक की **platform-independent** प्रकृति को दर्शाता है और ISAR प्रणालियों को अधिक लचीला बनाता है।

साथ ही, आधुनिक सूचना परिवेश में **interoperability** का महत्त्व बढ़ गया है, जिसके माध्यम से विभिन्न प्रणालियाँ और प्लेटफॉर्म परस्पर जुड़कर कार्य कर सकते हैं। इससे ISAR प्रणालियाँ अन्य डिजिटल लाइब्रेरी और रिपॉजिटरी के साथ प्रभावी सूचना-आदान-प्रदान कर पाती हैं।

उदाहरण (Concise Example)

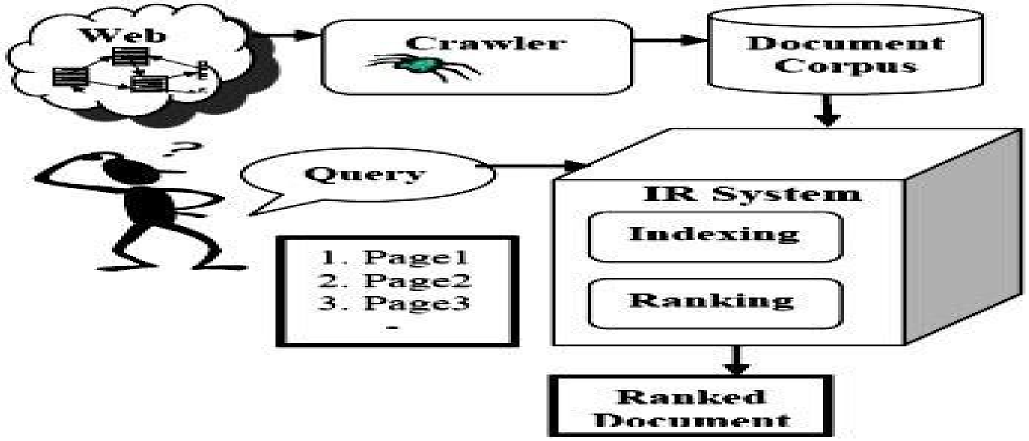
किसी विश्वविद्यालय की डिजिटल लाइब्रेरी पहले केवल **campus-based access** तक सीमित थी, जो तकनीकी परिवर्तन के बाद **web-based और mobile-friendly ISAR system** में परिवर्तित हो गई। इससे उपयोगकर्ता कहीं से भी ई-जर्नल, थीसिस और शोध-सामग्री तक पहुँच प्राप्त कर सकता है।

इस प्रकार, सूचना परिवेश एवं तकनीकी परिवर्तन ISAR प्रणालियों को **अधिक अनुकूलनशील, स्केलेबल और उपयोगकर्ता-उन्मुख** बनने के लिए प्रेरित करते हैं।



Information Retrieval का विकास (Evolution of Information Retrieval)

Explaination: यह चित्र दिखाता है कि जानकारी खोजने की प्रक्रिया 1960 के दशक की साधारण खोज से शुरू होकर **1990 के दशक में वेब खोज** और बाद में **संदर्भ के आधार पर स्मार्ट खोज (Contextual Search)** तक कैसे विकसित हुई।



<u>यह चित्र ISAR प्रणालियों के तकनीकी विकास को offline से web एवं AI-आधारित प्रणालियों तक दर्शाता है।</u>

11.6 ISAR प्रणालियों में लचीलापन (Flexibility in ISAR Systems):

ISAR प्रणालियों में **लचीलापन (Flexibility)** से तात्पर्य प्रणाली की उस क्षमता से है, जिसके द्वारा वह बिना पूर्ण पुनःडिज़ाइन के **नई सेवाओं, सूचना-प्रकारों और उपयोगकर्ता आवश्यकताओं** को समायोजित कर सके। लचीली ISAR प्रणाली बदलते सूचना परिवेश में निरंतर प्रभावी बनी रहती है।

लचीलापन इस बात पर निर्भर करता है कि ISAR प्रणाली अपने **खोज-तंत्र, सामग्री प्रबंधन और इंटरफेस** में कितनी आसानी से परिवर्तन स्वीकार कर सकती है। उदाहरणस्वरूप, नई सामग्री जोड़ना, खोज विकल्पों को विस्तृत करना या इंटरफेस को अद्यतन करना लचीली प्रणाली में सरल होता है।

ISAR प्रणालियों में लचीलापन मुख्यतः निम्न रूपों में देखा जा सकता है—

- **Search Flexibility**: simple एवं advanced search विकल्पों का समर्थन
- **Content Flexibility**: विभिन्न सूचना-प्रकारों को संभालने की क्षमता
- **Interface Flexibility**: उपयोगकर्ता के अनुसार इंटरफेस में परिवर्तन
- **System Flexibility**: नई सेवाओं या मॉड्यूल जोड़ने की सुविधा

11.7 Scalability एवं Interoperability (Scalability and Interoperability in ISAR Systems):

आधुनिक सूचना परिवेश में ISAR प्रणालियों से अपेक्षा की जाती है कि वे न केवल वर्तमान आवश्यकताओं को पूरा करें, बल्कि भविष्य में बढ़ने वाले **डेटा, उपयोगकर्ताओं और प्रणालियों**

के साथ भी प्रभावी बनी रहें। इस संदर्भ में **Scalability** और **Interoperability** ISAR प्रणालियों की दो अत्यंत महत्वपूर्ण विशेषताएँ हैं, जो प्रणाली की दीर्घकालिक उपयोगिता और प्रभावशीलता को सुनिश्चित करती हैं।

11.7.1 Scalability (स्केलेबिलिटी)

Scalability से आशय ISAR प्रणाली की उस क्षमता से है, जिसके द्वारा वह **डेटा की मात्रा, उपयोगकर्ताओं की संख्या या सेवाओं के विस्तार** के बावजूद अपने प्रदर्शन (performance) को बनाए रख सके।

सरल शब्दों में, यदि किसी ISAR प्रणाली में—

- रिकॉर्ड 10,000 से बढ़कर 10,00,000 हो जाएँ, या
- उपयोगकर्ता 100 से बढ़कर 10,000 हो जाएँ,

और फिर भी खोज की गति, सटीकता और सेवा-गुणवत्ता प्रभावित न हो, तो वह प्रणाली **scalable** कहलाती है।

Scalability के प्रमुख आयाम

- **Data Scalability** – बढ़ते संग्रह (books, journals, datasets, multimedia) को सँभालने की क्षमता
- **User Scalability** – एक साथ अधिक उपयोगकर्ताओं को सेवा प्रदान करना
- **Functional Scalability** – नई सेवाएँ या फीचर्स जोड़ने की क्षमता

उदाहरण

किसी विश्वविद्यालय की डिजिटल रिपॉजिटरी प्रारम्भ में केवल कुछ सौ थीसिस को सँभाल रही थी। समय के साथ जब उसमें हज़ारों ई-थीसिस, शोध-लेख और डेटासेट जुड़ गए, तब भी यदि प्रणाली बिना धीमी हुए कार्य करती रही, तो यह **Scalability** का स्पष्ट उदाहरण है।

11.7.2 Interoperability (इंटरऑपरेबिलिटी)

Interoperability से तात्पर्य ISAR प्रणाली की उस क्षमता से है, जिसके द्वारा वह **अन्य प्रणालियों, मानकों और प्लेटफॉर्म** के साथ सूचना का आदान-प्रदान कर सके।

अर्थात, विभिन्न डिजिटल लाइब्रेरी, OPAC, रिपॉजिटरी या सूचना प्रणालियाँ एक-दूसरे के साथ **समन्वय (coordination)** में कार्य कर सकें।

Interoperability के प्रमुख प्रकार

- **Technical Interoperability** – हार्डवेयर/सॉफ्टवेयर स्तर पर संगतता
- **Semantic Interoperability** – अर्थ एवं विषयवस्तु का समान बोध
- **Syntactic Interoperability** – डेटा संरचना एवं फॉर्मेट का सामंजस्य

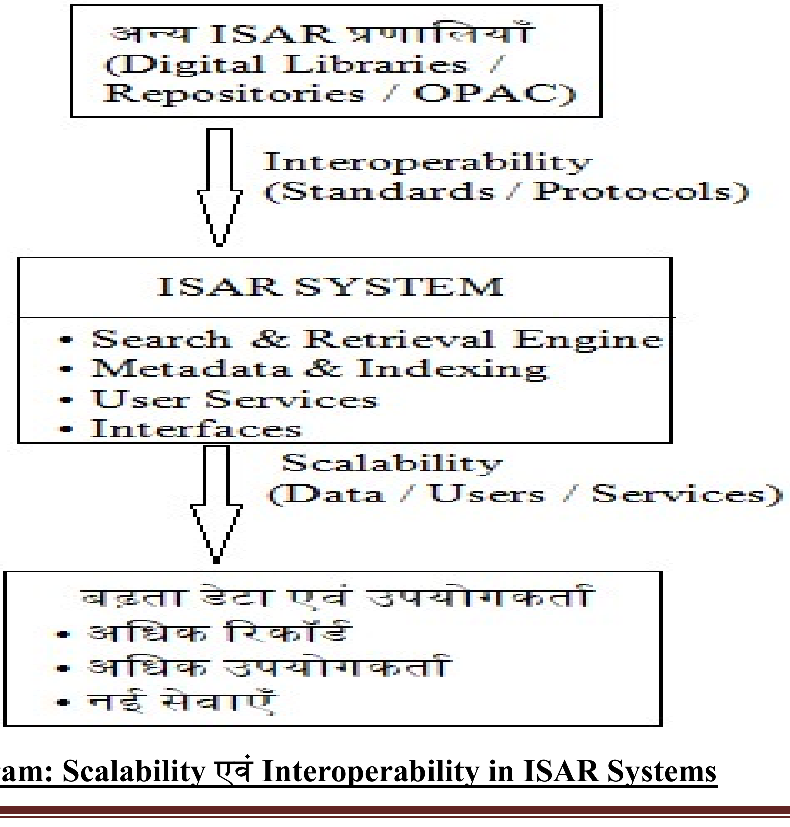
ISAR प्रणालियों में यह क्षमता सामान्यतः **मानकों एवं प्रोटोकॉल** (जैसे MARC, Dublin Core, Z39.50, OAI-PMH) के माध्यम से प्राप्त की जाती है।

उदाहरण

यदि एक पुस्तकालय का OPAC किसी अन्य विश्वविद्यालय की डिजिटल लाइब्रेरी से रिकॉर्ड खोज और प्राप्त कर सकता है, तो यह **Interoperability** का उदाहरण है।

Scalability एवं Interoperability का संयुक्त महत्व

Scalability ISAR प्रणाली को **आकार में बढ़ने योग्य** बनाती है, जबकि Interoperability उसे **अन्य प्रणालियों से जुड़ने योग्य** बनाती है। दोनों मिलकर ISAR प्रणाली को अधिक **लचीला, प्रभावी और भविष्य-उन्मुख** बनाते हैं।



<u>**Diagram: Scalability एवं Interoperability in ISAR Systems**</u>

चित्र: ISAR प्रणालियों में Scalability एवं Interoperability
यह चित्र दर्शाता है कि ISAR प्रणाली किस प्रकार बढ़ते डेटा, उपयोगकर्ताओं एवं सेवाओं को सँभालते हुए (Scalability) अन्य डिजिटल लाइब्रेरी और सूचना प्रणालियों के साथ मानकों एवं प्रोटोकॉल के माध्यम से समन्वय स्थापित करती है (Interoperability)।

Key Points :

- Scalability ISAR प्रणाली को बढ़ते डेटा और उपयोगकर्ताओं के लिए सक्षम बनाती है।
- Interoperability विभिन्न प्रणालियों के बीच सूचना-आदान-प्रदान को संभव बनाती है।
- दोनों अवधारणाएँ ISAR की **दीर्घकालिक स्थिरता (sustainability)** सुनिश्चित करती हैं।
- आधुनिक डिजिटल लाइब्रेरी एवं वेब-आधारित ISAR प्रणालियों के लिए ये अनिवार्य गुण हैं।

11.8 बहुभाषिक एवं बहु-विषयक अनुकूलन (Multilingual & Multidisciplinary Adaptation):

11.8.1 बहुभाषिक अनुकूलन (Multilingual Adaptation)

भारत जैसे बहुभाषी देश में उपयोगकर्ता हिंदी, अंग्रेज़ी तथा विभिन्न क्षेत्रीय भाषाओं में सूचना खोजते हैं। ऐसी स्थिति में ISAR प्रणालियों में **Cross-Language Information Retrieval (CLIR)** अत्यंत उपयोगी सिद्ध होता है।

CLIR का अर्थ है—जब उपयोगकर्ता की खोज-भाषा (query language) और दस्तावेज़ की भाषा अलग-अलग हो, तब भी प्रासंगिक सूचना की पुनर्प्राप्ति संभव हो। यह सुविधा **भाषा-मैपिंग, अनुवाद तथा नियंत्रित शब्दावली (thesaurus)** के माध्यम से प्राप्त की जाती है।

उदाहरण (CLIR): यदि उपयोगकर्ता हिंदी में "जलवायु परिवर्तन" खोजता है, तो ISAR प्रणाली अंग्रेज़ी भाषा में उपलब्ध "Climate Change" से संबंधित दस्तावेज़ भी पुनर्प्राप्त कर सकती है।

11.8.2 बहु-विषयक अनुकूलन (Multidisciplinary Adaptation)

आधुनिक शोध कार्य प्रायः **interdisciplinary** होता है, जिसमें एक ही विषय से संबंधित सूचना विभिन्न विषय-क्षेत्रों में फैली होती है। इसलिए ISAR प्रणालियों में **metadata, indexing तथा vocabularies** को विभिन्न विषयों के बीच **cross-domain mapping** के लिए अनुकूलित करना आवश्यक होता है।

उदाहरणस्वरूप, "पर्यावरण" विषय से संबंधित खोज में **अर्थशास्त्र, सार्वजनिक नीति एवं सामाजिक विज्ञान** से जुड़े संसाधनों का एकीकृत रूप में प्राप्त होना बहु-विषयक अनुकूलन का संकेत है।

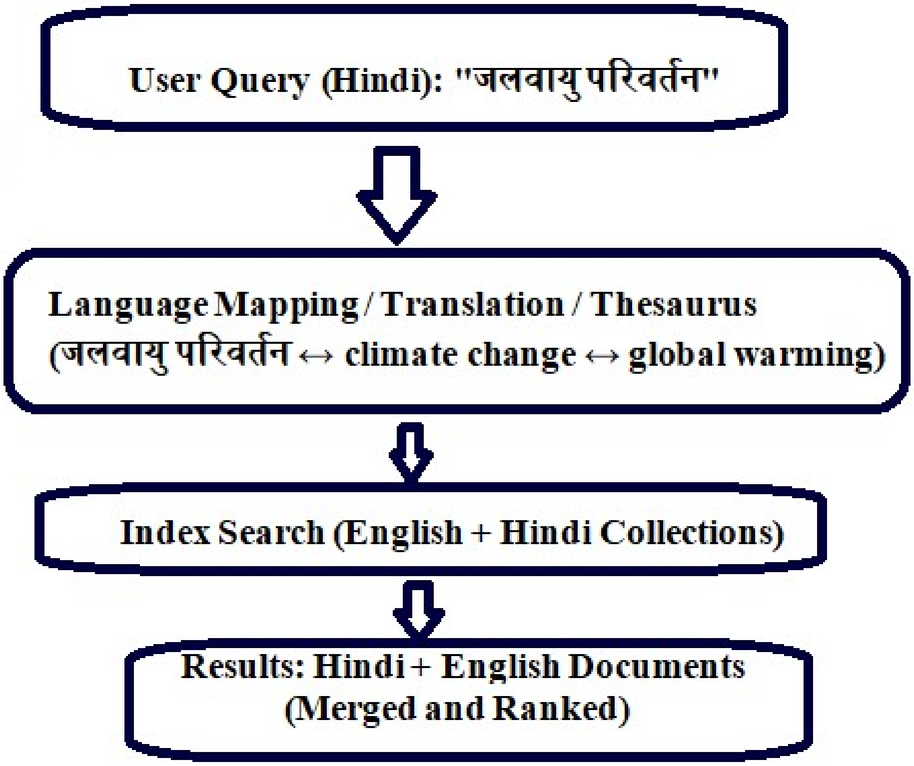


Diagram: CLIR Concept

यह चित्र Cross-Language Information Retrieval (CLIR) की कार्यप्रणाली को दर्शाता है, जिसमें एक भाषा की query के माध्यम से अन्य भाषाओं में उपलब्ध प्रासंगिक सूचना पुनर्प्राप्त की जाती है।

11.9 डिजिटल रिपॉजिटरी एवं वेब वातावरण में अनुकूलन (Adaptation in Digital Repository and Web Environment):

डिजिटल रिपॉजिटरी एवं वेब वातावरण में ISAR प्रणालियों की अनुकूलन क्षमता का अर्थ है—
प्रणाली का **डिजिटल मानकों, वेब तकनीकों और नेटवर्क-आधारित सेवाओं** के अनुरूप स्वयं को ढालना, ताकि सूचना की खोज, पहुँच और पुनर्प्राप्ति अधिक प्रभावी हो सके।

डिजिटल रिपॉजिटरी में अनुकूलन मुख्यतः निम्न पहलुओं में दिखाई देता है—

- **Metadata Standards**: Dublin Core जैसे मानकों का उपयोग, जिससे संसाधनों का विवरण मानकीकृत और interoperable बनता है।
- **Harvesting Mechanisms**: OAI-PMH के माध्यम से विभिन्न रिपॉजिटरी से मेटाडेटा का संग्रह और साझा करना।
- **Object Linking**: खोज परिणामों से मूल डिजिटल ऑब्जेक्ट (full-text, thesis, dataset) तक सीधी पहुँच।
- **Semantic Enrichment**: खोज इंजनों और वेब उपयोगकर्ताओं के लिए विषय-वस्तु को अधिक अर्थपूर्ण (context-aware) बनाना।

वेब वातावरण में यह अनुकूलन ISAR प्रणालियों को **search-engine friendly**, **globally accessible** और **integrated digital knowledge systems** के रूप में विकसित करता है।

11.10 उपयोगकर्ता-केंद्रित ISAR प्रणालियाँ (User-Centred ISAR Systems):

उपयोगकर्ता-केंद्रित ISAR प्रणालियाँ वे प्रणालियाँ हैं, जिनका डिज़ाइन एवं संचालन **उपयोगकर्ता की सूचना आवश्यकताओं, खोज व्यवहार और सुविधा** को केंद्र में रखकर किया जाता है। ऐसी प्रणालियाँ उपयोगकर्ता अनुभव को सरल, प्रभावी और वैयक्तिकृत बनाती हैं।

User-centred adaptability के प्रमुख practical features—

1. **Simple & Intuitive Interface**: आसान नेविगेशन और स्पष्ट खोज विकल्प
2. **Personalisation**: user profile के आधार पर recommendations एवं saved searches
3. **Adaptive Search Support**: query suggestions, filters और relevance ranking
4. **Feedback Mechanism**: user feedback एवं search logs के आधार पर सुधार
5. **Accessibility**: मोबाइल-फ्रेंडली डिज़ाइन एवं विविध उपयोगकर्ताओं के लिए सुविधा

6. **User friendliness**: ease of use आवश्यक है

11.11 अनुकूलन से संबंधित चुनौतियाँ (Challenges in Adaptability of ISAR Systems):

ISAR प्रणालियों में अनुकूलन क्षमता विकसित करना आवश्यक है, किंतु इसके साथ अनेक **तकनीकी, संगठनात्मक और नैतिक चुनौतियाँ** भी जुड़ी होती हैं। प्रमुख चुनौतियाँ निम्नलिखित हैं—

1. **Cost एवं Infrastructure**: उन्नत सर्वर, क्लाउड सेवाएँ, नेटवर्क एवं पर्याप्त bandwidth की आवश्यकता होती है, जिससे संस्थानों पर वित्तीय दबाव बढ़ता है।
2. **Standards की जटिलता**: MARC, Dublin Core, OAI-PMH, Z39.50 जैसे अनेक मानकों को समझना और एकीकृत करना तकनीकी रूप से चुनौतीपूर्ण होता है।
3. **Data Quality Issues**: असंगत या अपूर्ण metadata, duplicate records और indexing errors खोज की गुणवत्ता को प्रभावित करते हैं।
4. **Privacy एवं Ethical Concerns**: user logs, search behaviour और personalization से संबंधित डेटा के संरक्षण एवं नैतिक उपयोग की चुनौती बनी रहती है।
5. **Human Resource एवं Training**: ISAR प्रणालियों के प्रभावी संचालन हेतु प्रशिक्षित मानव संसाधन की आवश्यकता होती है, जबकि कई संस्थानों में skill gap पाया जाता है।
6. **Change Management**: नए इंटरफेस या प्रणाली में परिवर्तन के प्रति उपयोगकर्ताओं एवं कर्मचारियों का प्रतिरोध (resistance) अनुकूलन प्रक्रिया को धीमा कर सकता है।

11.12 व्यावहारिक उदाहरण : Adaptive ISAR System (Case Model)

Case: *University Repository + OPAC Integration*

स्थिति (Situation)

विश्वविद्यालय में पुस्तकालय का **OPAC (books)** और **Digital Repository (thesis/articles)** अलग-अलग प्रणालियों में संचालित हैं। उपयोगकर्ताओं की अपेक्षा है कि उन्हें **single search box** के माध्यम से सभी संसाधनों तक पहुँच प्राप्त हो।

Adaptive Solution (अनुकूलन आधारित समाधान)

इस आवश्यकता को पूरा करने के लिए निम्न **adaptive features** अपनाए जाते हैं—

- **Unified Search Interface** का विकास
- **Repository metadata harvesting** द्वारा *OAI-PMH* का उपयोग
- आवश्यकतानुसार **bibliographic interoperability** (standards-based exchange)
- **Result merging** एवं **duplicate removal**
- बहुभाषी खोज हेतु **Multilingual CLIR module**

Expected Output (अपेक्षित परिणाम) : इन अनुकूलनों के पश्चात उपयोगकर्ता को **one-click access** के माध्यम से—

- थीसिस का **full-text PDF**
- संबंधित पुस्तक का **OPAC record**
- विषय-संबंधित **research article links**

एकीकृत एवं ranked रूप में प्राप्त होते हैं।

इस प्रकार, यह केस मॉडल दर्शाता है कि एक **Adaptive ISAR System** विभिन्न सूचना प्रणालियों को integrate कर **user-centred, efficient और seamless information retrieval** सुनिश्चित करता है।

11.13 आधुनिक प्रवृत्तियाँ एवं भविष्य की दिशा (Modern Trends and Future Directions in ISAR Systems):

आधुनिक ISAR प्रणालियाँ पारंपरिक keyword-based search से आगे बढ़कर **intelligent, interactive और user-centric systems** की ओर विकसित हो रही हैं। ISAR से जुड़ी प्रमुख आधुनिक प्रवृत्तियाँ एवं भविष्य की दिशाएँ निम्नलिखित हैं—

1. **AI-based Semantic Search:** AI एवं semantic technologies के माध्यम से ISAR प्रणालियाँ शब्दों के बजाय **concept-level retrieval** प्रदान करती हैं, जिससे अधिक प्रासंगिक परिणाम प्राप्त होते हैं।
2. **Question Answering Systems:** Natural Language Processing (NLP) पर आधारित प्रणालियाँ उपयोगकर्ता के प्रश्नों के **direct answers** प्रदान करने में सक्षम होती जा रही हैं।

3. **Voice Search एवं Conversational OPAC:** Voice-enabled एवं conversational interfaces उपयोगकर्ताओं को बोलकर खोज करने की सुविधा प्रदान करते हैं, जो विशेष रूप से mobile एवं accessibility के लिए उपयोगी हैं।
4. **Linked Data एवं Knowledge Graph:** Linked Data और Knowledge Graph के उपयोग से सूचना संसाधनों के बीच **semantic linking** संभव होती है, जिससे context-aware retrieval में सुधार होता है।
5. **Predictive Recommendation Systems:** User behaviour और search patterns के आधार पर ISAR प्रणालियाँ **personalised एवं predictive recommendations** प्रदान करती हैं।

11.14 सारांश (Summary):

इस इकाई में ISAR प्रणालियों की **अनुकूलन क्षमता (Adaptability)** का अध्ययन किया गया। बदलते सूचना परिवेश, तकनीकी विकास और उपयोगकर्ता व्यवहार के कारण आधुनिक ISAR प्रणालियों का **लचीला, स्केलेबल और इंटरऑपरेबल** होना आवश्यक है।

इकाई में यह स्पष्ट किया गया कि **बहुभाषिक एवं बहु-विषयक अनुकूलन**, **डिजिटल रिपॉजिटरी एवं वेब-आधारित वातावरण**, तथा **उपयोगकर्ता-केंद्रित दृष्टिकोण** ISAR प्रणालियों की प्रभावशीलता को बढ़ाते हैं। साथ ही, अनुकूलन से जुड़ी प्रमुख **तकनीकी, संगठनात्मक और नैतिक चुनौतियों** पर भी चर्चा की गई।

अंततः, **AI-आधारित खोज, voice search, linked data और predictive recommendation** जैसी आधुनिक प्रवृत्तियाँ ISAR प्रणालियों को अधिक **intelligent, adaptive और future-ready** बना रही हैं।

11.15 शब्दावली (Glossary)

- **Adaptability (अनुकूलन क्षमता):** ISAR प्रणाली की वह क्षमता जिसके द्वारा वह बदलती तकनीक, उपयोगकर्ता आवश्यकताओं और सूचना परिवेश के अनुरूप स्वयं को समायोजित कर सके।
- **Flexibility (लचीलापन):** बिना core structure बदले नई सुविधाएँ, सेवाएँ या मॉड्यूल जोड़ने अथवा संशोधित करने की क्षमता।
- **Scalability (स्केलेबिलिटी):** ISAR प्रणाली की वह क्षमता जिसके द्वारा वह बढ़ते डेटा, उपयोगकर्ताओं एवं सेवाओं के बावजूद कुशल प्रदर्शन बनाए रख सके।

- **Interoperability (इंटरऑपरेबिलिटी):** विभिन्न सूचना प्रणालियों का मानकों एवं प्रोटोकॉल के माध्यम से आपस में समन्वय कर सूचना का आदान-प्रदान करना।
- **User-Centred ISAR:** वह ISAR प्रणाली जिसका डिज़ाइन और सेवाएँ उपयोगकर्ता की सूचना आवश्यकताओं और खोज व्यवहार पर आधारित हों।
- **Multilingual Adaptation (बहुभाषिक अनुकूलन):** एक से अधिक भाषाओं में सूचना के भंडारण, खोज एवं पुनर्प्राप्ति की क्षमता।
- **Multidisciplinary Adaptation (बहु-विषयक अनुकूलन):** विभिन्न विषय-क्षेत्रों से संबंधित सूचना को एकीकृत रूप में संगठित एवं पुनर्प्राप्त करने की क्षमता।
- **CLIR (Cross-Language Information Retrieval):** ऐसी सूचना पुनर्प्राप्ति प्रक्रिया जिसमें उपयोगकर्ता की query और दस्तावेज़ की भाषा अलग होने पर भी खोज संभव हो।
- **Digital Repository (डिजिटल रिपॉजिटरी):** शोध-थीसिस, लेख, डेटासेट आदि डिजिटल संसाधनों का संगठित भंडार।
- **Metadata (मेटाडेटा):** सूचना संसाधन के बारे में विवरणात्मक जानकारी, जो खोज और संगठन में सहायक होती है।
- **OAI-PMH:** डिजिटल रिपॉजिटरी के बीच मेटाडेटा harvesting हेतु प्रयुक्त मानक प्रोटोकॉल।
- **Semantic Search:** शब्दों के बजाय उनके अर्थ (concept) के आधार पर सूचना खोजने की तकनीक।
- **Knowledge Graph:** सूचना संसाधनों के बीच semantic relationships को दर्शाने वाली संरचना।
- **Predictive Recommendation:** उपयोगकर्ता व्यवहार के आधार पर भविष्य की सूचना आवश्यकताओं का अनुमान लगाकर सुझाव प्रदान करना।

1.16 निबंधात्मक प्रश्न (Essay Type Questions) – with Answers

प्रश्न 1. ISAR प्रणालियों में **अनुकूलन क्षमता (Adaptability)** की अवधारणा को स्पष्ट कीजिए।

उत्तर: ISAR प्रणालियों में अनुकूलन क्षमता से आशय प्रणाली की उस योग्यता से है, जिसके माध्यम से वह बदलती तकनीक, उपयोगकर्ता आवश्यकताओं, सूचना संसाधनों और कार्य-परिवेश के अनुरूप स्वयं को ढाल सके। आधुनिक सूचना परिवेश स्थिर न होकर निरंतर परिवर्तनशील है, इसलिए ISAR प्रणालियों का adaptable होना अनिवार्य है। Adaptability

के अंतर्गत Flexibility, Scalability और Interoperability जैसे घटक आते हैं, जो प्रणाली को दीर्घकालिक रूप से प्रभावी और उपयोगी बनाए रखते हैं।

प्रश्न 2. ISAR प्रणालियों में **Flexibility, Scalability और Interoperability** के महत्व की विवेचना कीजिए।

उत्तर: Flexibility ISAR प्रणाली को बिना मूल संरचना बदले नई सेवाएँ और फीचर्स जोड़ने में सक्षम बनाती है। Scalability बढ़ते डेटा और उपयोगकर्ताओं के बावजूद प्रणाली के प्रदर्शन को बनाए रखती है। Interoperability विभिन्न सूचना प्रणालियों के बीच समन्वय स्थापित कर सूचना के आदान-प्रदान को संभव बनाती है। इन तीनों का संयुक्त प्रभाव ISAR प्रणालियों को भविष्य-उन्मुख और टिकाऊ बनाता है।

प्रश्न 3. उपयोगकर्ता आवश्यकताओं एवं व्यवहार का ISAR प्रणालियों के डिज़ाइन पर क्या प्रभाव पड़ता है?

उत्तर: आधुनिक उपयोगकर्ता तेज़, सरल और वैयक्तिकृत सूचना सेवाएँ चाहता है। इसी कारण ISAR प्रणालियाँ user-centred design, personalised search, filters, alerts और feedback mechanisms अपनाती हैं। उपयोगकर्ता व्यवहार के विश्लेषण से खोज परिणामों की प्रासंगिकता और उपयोगिता में सुधार होता है।

प्रश्न 4.ISAR प्रणालियों में **बहुभाषिक एवं बहु-विषयक अनुकूलन** का महत्व स्पष्ट कीजिए।

उत्तर:
बहुभाषिक अनुकूलन विभिन्न भाषाओं में खोज एवं पुनर्प्राप्ति को संभव बनाता है, जबकि बहु-विषयक अनुकूलन विभिन्न विषय-क्षेत्रों की सूचना को एकीकृत करता है। CLIR, multilingual metadata और cross-domain vocabularies के माध्यम से ISAR प्रणालियाँ अधिक समावेशी और व्यापक बनती हैं।

प्रश्न 5.डिजिटल रिपॉजिटरी एवं वेब वातावरण में ISAR प्रणालियों के अनुकूलन की भूमिका समझाइए।

उत्तर:
डिजिटल रिपॉजिटरी में ISAR प्रणालियाँ metadata standards, OAI-PMH harvesting, object linking और semantic enrichment के माध्यम से अनुकूलन प्राप्त करती हैं। इससे सूचना संसाधन globally accessible, interoperable और search-engine friendly बनते हैं।

प्रश्न 6. ISAR प्रणालियों में अनुकूलन से संबंधित प्रमुख चुनौतियों का वर्णन कीजिए।

उत्तर: प्रमुख चुनौतियों में उच्च लागत, infrastructure की आवश्यकता, standards की जटिलता, metadata की गुणवत्ता, privacy एवं ethical concerns, प्रशिक्षित मानव संसाधन की कमी तथा change management शामिल हैं। इन चुनौतियों के प्रभावी प्रबंधन के बिना सफल अनुकूलन संभव नहीं है।

प्रश्न 7. Adaptive ISAR System को उदाहरण सहित स्पष्ट कीजिए।

उत्तर: जब OPAC और digital repository को unified search interface के माध्यम से एकीकृत किया जाता है, metadata harvesting और result merging अपनाया जाता है, तब उपयोगकर्ता को एक ही खोज में पुस्तक, थीसिस और शोध-लेख उपलब्ध हो जाते हैं। यह Adaptive ISAR System का व्यावहारिक उदाहरण है।

प्रश्न 8. ISAR प्रणालियों की आधुनिक प्रवृत्तियों एवं भविष्य की दिशा पर प्रकाश डालिए।

उत्तर: AI-based semantic search, question answering systems, voice search, linked data, knowledge graph और predictive recommendation जैसी प्रवृत्तियाँ ISAR प्रणालियों को अधिक intelligent, interactive और user-centric बना रही हैं।

11.17 लघु प्रश्न (Short Answer Questions)

1. ISAR प्रणाली से आप क्या समझते हैं?
2. Adaptability और Flexibility में अंतर लिखिए।
3. Scalability का क्या महत्व है?
4. Interoperability से आप क्या समझते हैं?
5. CLIR का संक्षिप्त परिचय दीजिए।
6. User-centred ISAR क्या है?
7. Metadata का ISAR में क्या योगदान है?
8. OAI-PMH का उपयोग क्यों आवश्यक है?

11.17.1 बहुविकल्पीय प्रश्न (MCQs) (प्रत्येक प्रश्न का एक सही उत्तर है)

1. **Adaptability का सर्वाधिक उपयुक्त अर्थ है—**
   a) केवल storage बढ़ाना
   b) बदलती जरूरतों के अनुसार सिस्टम का समायोजन ✔

c) केवल UI बदलना
d) केवल hardware बदलना

2. **ISAR system का front-end क्या कहलाता है?**
a) Knowledge base
b) User Interface ✔
c) Index
d) Server

3. **Query Interface का मुख्य कार्य क्या है?**
a) केवल report printing
b) Query input लेना तथा उसे refine/translate करना ✔
c) Shelving
d) Binding

4. **Advanced search में सामान्यतः किसका प्रयोग होता है?**
a) Barcode
b) Boolean operators ✔
c) RFID only
d) None

5. **Scalability का सम्बन्ध किससे है?**
a) भाषा से
b) System capacity बढ़ने पर performance से ✔
c) Cataloguing rules से
d) Photocopy से

6. **Interoperability का अर्थ है—**
a) केवल offline access
b) विभिन्न प्रणालियों का साथ मिलकर कार्य करना ✔
c) Index हटाना
d) User registration

7. **Z39.50 किसके लिए जाना जाता है?**
a) Binding standard
b) Search एवं retrieval protocol ✔
c) Printing code
d) MARC replacement

8. **Dublin Core का उपयोग किसके description के लिए होता है?**
a) Web resources ✔
b) Cement bags

c) Hardware drivers
d) Attendance

9. **OAI-PMH का मुख्य कार्य क्या है?**
a) Book repair
b) Metadata harvesting ✔
c) Shelving
d) Gate entry

10. **CLIR (Cross-Language Information Retrieval) का उपयुक्त उदाहरण है—**
a) English query → English documents only
b) Hindi query → English documents retrieve ✔
c) Hindi query → Hindi documents only
d) Voice search only

11. Scalability का मुख्य संबंध किससे है?
a) सुरक्षा
b) विस्तार-योग्यता ✔
c) वर्गीकरण
d) सूचीकरण

➢ Match the Following

Match List–A with List–B

List–A	**List–B**
Z39.50	Cross-system search
Dublin Core	Web resource description
OAI-PMH	Metadata harvesting
Query Interface	Query refine एवं translation
Scalability	Big data / user load performance

➢ Fill in the Blanks

1. ISAR system का __________ उपयोगकर्ताओं के साथ interaction कराता है।
उत्तर: User Interface

2. OAI-PMH का उपयोग __________ के लिए किया जाता है।
उत्तर: Metadata harvesting

11.18 अभ्यास प्रश्नों के उत्तर (Indicative Answers)

संकेत:

- Adaptability = बदलते परिवेश के अनुरूप प्रणाली का समायोजन
- Flexibility = फीचर्स/सेवाओं में परिवर्तन की सुविधा
- Scalability = बढ़ते डेटा/यूज़र पर भी कुशल प्रदर्शन
- Interoperability = विभिन्न प्रणालियों का आपसी समन्वय
- CLIR = अलग भाषा में भी सूचना पुनर्प्राप्ति

11.19 संदर्भ ग्रंथ सूची (References)

1. गोस्वामी, जे. एन. – सूचना भंडारण एवं पुनर्प्राप्ति प्रणाली
2. चतुर्वेदी, एस. – डिजिटल पुस्तकालय एवं सूचना सेवाएँ
3. शर्मा, आर. – आधुनिक पुस्तकालय विज्ञान
4. IGNOU – Information Storage and Retrieval Systems (SLM)
5. यूजीसी-DEB अध्ययन सामग्री – ISAR Systems
6. भारतीय पुस्तकालय संघ (ILA) – प्रकाशन
7. सिंह, के. पी. – सूचना संगठन एवं पुनर्प्राप्ति
8. वर्मा, पी. – डिजिटल सूचना संसाधन
9. राष्ट्रीय पुस्तकालय, कोलकाता – प्रशिक्षण सामग्री
10. यूजीसी-स्वयं/नैप्टेल – LIS पाठ्यसामग्री
11. Chowdhury, G. G. – Introduction to Modern Information Retrieval
12. Baeza-Yates, R. & Ribeiro-Neto, B. – Modern Information Retrieval
13. Arms, W. Y. – Digital Libraries
14. Rowley, J. – Information Retrieval
15. Borgman, C. L. – From Gutenberg to the Global Information Infrastructure
16. Tefko Saracevic – Selected writings on IR
17. Witten, I. H. et al. – Managing Gigabytes
18. ISO / IFLA documentation on metadata & interoperability
19. Dublin Core Metadata Initiative – Official Documentation

20. Chowdhury, G. G. *Introduction to Modern Information Retrieval*.

21. Lancaster, F. W. *Information Retrieval Systems: Characteristics, Testing and Evaluation*.

22. Manning, C. D., Raghavan, P., & Schütze, H. *Introduction to Information Retrieval*.

चतुर्थ खण्ड

सूचना पुनर्प्राप्ति

(Information Retrieval)

इकाई–12 : बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियाँ (Intelligent Information Retrieval Systems)

इकाई की रूपरेखा

12.1 प्रस्तावना

12.2 उद्देश्य

12.3 बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली: अर्थ एवं अवधारणा

12.4 पारंपरिक IR बनाम बुद्धिमान IR प्रणाली

12.5 Artificial Intelligence (AI) का IR में प्रयोग

12.6 Machine Learning एवं Information Retrieval

12.7 Natural Language Processing (NLP) और IR

12.8 Semantic Search एवं Context-based Retrieval

12.9 Recommender Systems एवं Personalization

12.10 Intelligent Search Engines

12.11 बुद्धिमान IR प्रणालियों के लाभ

12.12 सीमाएँ एवं चुनौतियाँ

12.13 व्यावहारिक उदाहरण: Intelligent IR system

12.14 भविष्य की प्रवृत्तियाँ

12.15 सारांश

12.16 शब्दावली

12.17 निबंधात्मक प्रश्न

12.18 लघु प्रश्न एवं MCQ

12.19 अभ्यास प्रश्नों के उत्तर

12.20 संदर्भ ग्रंथ सूची

12.1 प्रस्तावना (Introduction)

सूचना विस्फोट (Information Explosion) के वर्तमान युग में केवल सूचना का संग्रह करना पर्याप्त नहीं रह गया है; बल्कि **उपयोगकर्ता की सूचना आवश्यकता के अनुरूप, सही समय पर और प्रासंगिक सूचना उपलब्ध कराना** अत्यंत आवश्यक हो गया है। पारंपरिक सूचना पुनर्प्राप्ति (Information Retrieval – IR) प्रणालियाँ मुख्यतः **कीवर्ड आधारित खोज** पर आधारित थीं, जो आज के **संदर्भ-आधारित, बहुभाषी और जटिल सूचना परिवेश** में सीमित सिद्ध होती हैं।

इन्हीं सीमाओं के समाधान के रूप में **बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियों (Intelligent Information Retrieval Systems – IIRS)** का विकास हुआ है। ये प्रणालियाँ **Artificial Intelligence, Machine Learning, Natural Language Processing तथा Semantic Technologies** का उपयोग करके सूचना के **अर्थ, संदर्भ और उपयोगकर्ता व्यवहार** को समझती हैं तथा **अधिक सटीक, प्रासंगिक और व्यक्तिगत खोज परिणाम** प्रदान करती हैं।

12.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात विद्यार्थी—

- बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियों की **अवधारणा एवं मूलभूत सिद्धांतों** को समझ सकेंगे।
- पारंपरिक सूचना पुनर्प्राप्ति प्रणाली और बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली के बीच **अंतर स्पष्ट कर सकेंगे।**
- सूचना पुनर्प्राप्ति में **Artificial Intelligence (AI), Machine Learning (ML)** एवं **Natural Language Processing (NLP)** की भूमिका का **विश्लेषण कर सकेंगे।**
- **Semantic Search तथा Recommender Systems** की कार्यप्रणाली को समझ सकेंगे।
- **आधुनिक Intelligent Search Engines** के कार्य एवं विशेषताओं का अध्ययन कर सकेंगे।
- बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियों से संबंधित **सीमाओं, चुनौतियों तथा भविष्य की प्रवृत्तियों का मूल्यांकन कर सकेंगे।**

12.3 बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली: अर्थ एवं अवधारणा (Meaning and Concept of Intelligent Information Retrieval System)

बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली (Intelligent Information Retrieval System – IIRS) वह प्रणाली है जो केवल **keywords matching** पर निर्भर नहीं रहती, बल्कि सूचना खोज (information search) के दौरान निम्नलिखित पहलुओं को ध्यान में रखती है—

- **Meaning (अर्थ)** – शब्दों का वास्तविक अर्थ
- **Context (संदर्भ)** – खोज किस स्थिति या विषय से जुड़ी है
- **Information Need (सूचना आवश्यकता)** – उपयोगकर्ता वास्तव में क्या जानना चाहता है
- **User Behaviour (उपयोगकर्ता व्यवहार)** – पूर्व खोज, रुचियाँ, प्राथमिकताएँ

इस प्रकार, IIRS सूचना को केवल *retrieve* नहीं करती, बल्कि **understand, analyse और personalise** भी करती है।

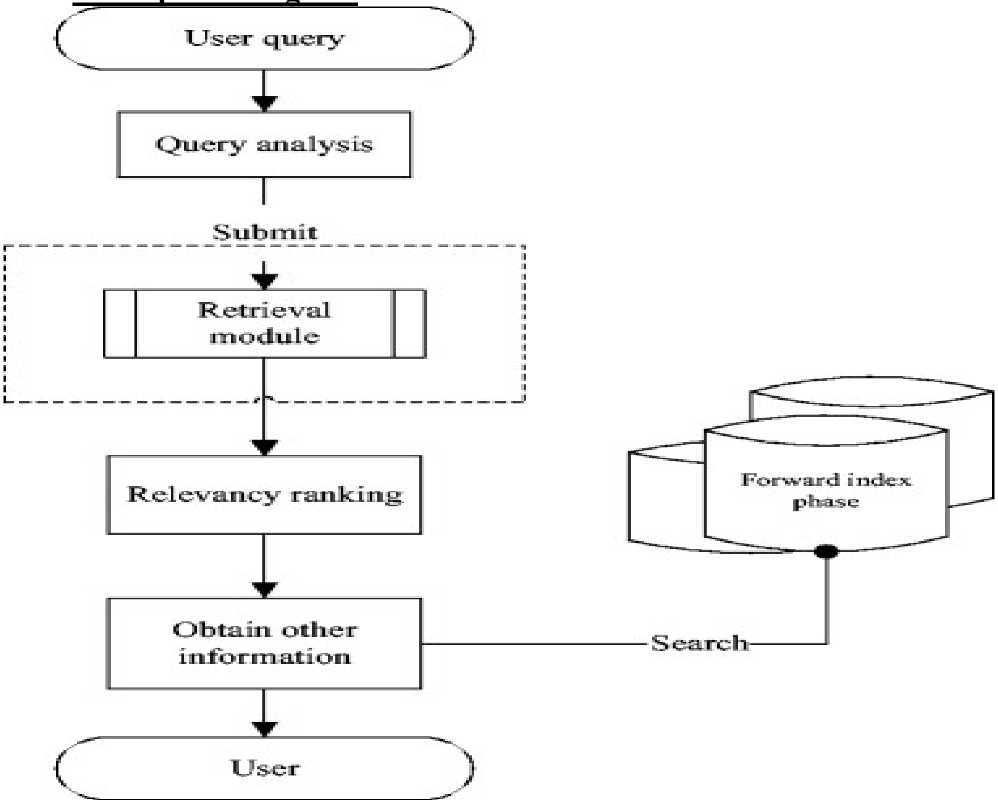
English Definition : "An Intelligent Information Retrieval System is a system that uses artificial intelligence techniques to understand the meaning, context, and user intent behind a query in order to retrieve more relevant, accurate, and personalized information."

उदाहरण (Example): यदि उपयोगकर्ता खोजता है—
"Apple"

- पारंपरिक IR → Apple (फल) + Apple (कंपनी) दोनों दिखा देगा
- बुद्धिमान IR → उपयोगकर्ता के **context** के आधार पर तय करेगा कि
  - ➢ Technology → *Apple Inc.*
  - ➢ Nutrition → *Apple (fruit)*

User Query
⇩
Understanding Meaning & Context
⇩
AI + ML + NLP Processing
⇩
User Profile & Knowledge Base
⇩
Relevant & Personalized Results

Conceptual Diagram



यह diagram concept clear करने के लिए है

12.4 पारंपरिक IR बनाम बुद्धिमान IR प्रणाली (Traditional IR vs Intelligent IR Systems):

Basis	Traditional IR	Intelligent IR
Search approach	Keyword-based	Meaning & Context-based
User understanding	Not considered	Considered (User profiling)
Personalization	No	Yes
Language processing	Limited	Advanced (NLP based)
Learning ability	Absent	Present (Machine Learning)
Search results	Same for all users	Personalized results

Key Difference :(for Exam)

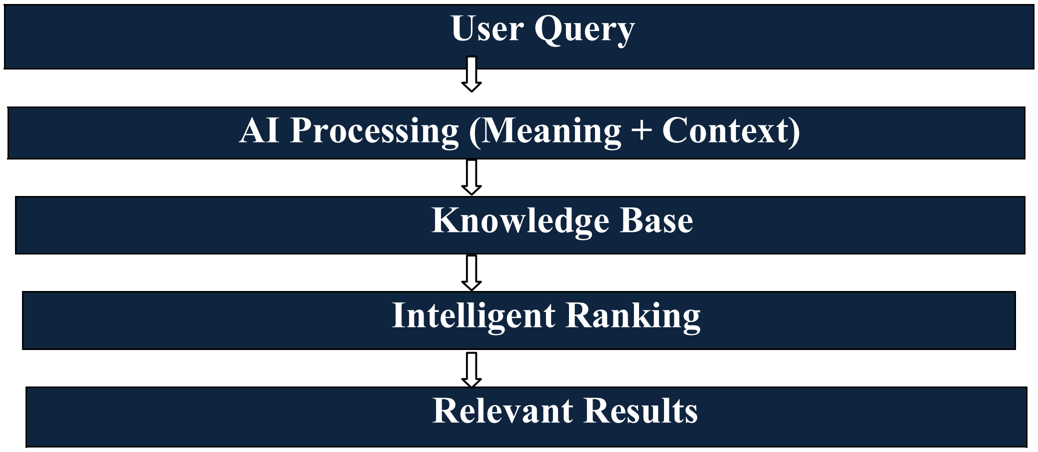
Traditional IR focuses on "what is written", whereas Intelligent IR focuses on "what is meant".

12.5 Artificial Intelligence (AI) का Information Retrieval (IR) में प्रयोग:

Artificial Intelligence (AI) ने सूचना पुनर्प्राप्ति प्रणालियों को **static (स्थिर)** से **dynamic (गतिशील)** बना दिया है। पारंपरिक IR प्रणालियाँ जहाँ केवल शब्दों के मिलान (keyword matching) पर आधारित थीं, वहीं AI आधारित IR प्रणालियाँ **meaning, context और user intent** को समझकर खोज परिणाम प्रदान करती हैं।

AI के प्रमुख योगदान (Major Contributions of AI in IR)

- **Knowledge Representation** – ज्ञान को संरचित रूप में प्रस्तुत करना
- **Inference & Reasoning** – तर्क के आधार पर निष्कर्ष निकालना
- **User Behaviour Analysis** – उपयोगकर्ता के खोज व्यवहार का विश्लेषण
- **Intelligent Ranking** – दस्तावेज़ों को प्रासंगिकता के अनुसार क्रमबद्ध करना
-



IR System Flow (Conceptual Diagram)

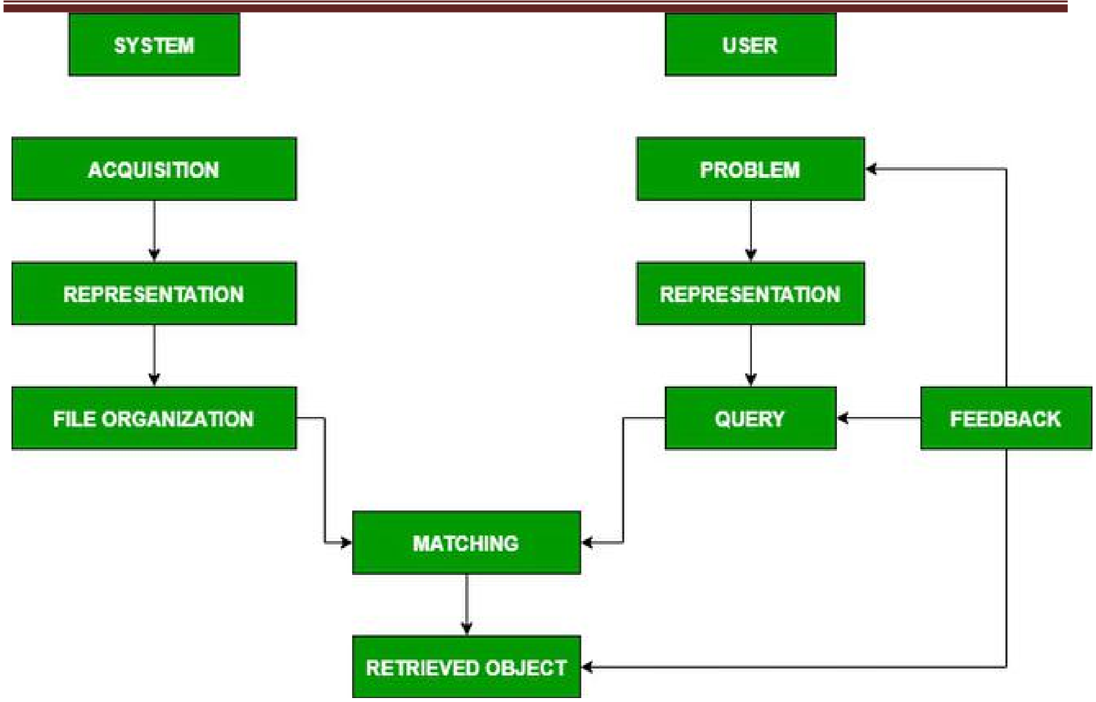


Diagram: Components of Information Retrieval

Exam-oriented point:

AI enables IR systems to *understand* information needs rather than merely *match* keywords.

12.6 Machine Learning (ML) एवं Information Retrieval

Machine Learning (ML) सूचना पुनर्प्राप्ति प्रणाली को **experience से सीखने (learning from data)** की क्षमता प्रदान करता है। ML आधारित IR प्रणालियाँ उपयोगकर्ता के **past searches, feedback और interaction patterns** के आधार पर समय के साथ अपनी कार्यक्षमता में सुधार करती हैं।

IR में Machine Learning के प्रमुख उपयोग

- **Learning to Rank** – खोज परिणामों की रैंकिंग को स्वतः सुधारना
- **Query Expansion** – समानार्थी (synonyms) या संबंधित शब्द जोड़ना
- **Spam Filtering** – अप्रासंगिक या अवांछित दस्तावेज़ों को हटाना
- **Document Classification** – दस्तावेज़ों को विषयवार वर्गीकृत करना

उदाहरण (Example)

यदि कोई उपयोगकर्ता बार-बार **"Digital Libraries"** विषय से संबंधित खोज करता है, तो ML आधारित IR प्रणाली भविष्य में उसी विषय से संबंधित दस्तावेज़ों, लेखों और संसाधनों को **higher priority (अधिक प्राथमिकता)** देगी।

12.7 Natural Language Processing (NLP) और Information Retrieval

Natural Language Processing (NLP) वह तकनीक है जो मशीन को **मानव भाषा (Human Language)** को समझने, विश्लेषण करने और संसाधित करने में सक्षम बनाती है। सूचना पुनर्प्राप्ति प्रणालियों में NLP के माध्यम से **user query का linguistic analysis** किया जाता है, जिससे खोज परिणाम अधिक सटीक और अर्थपूर्ण बनते हैं।

NLP के प्रमुख घटक (Key Components of NLP)

- **Tokenization** – वाक्य को शब्दों या tokens में विभाजित करना
- **Stemming / Lemmatization** – शब्दों को उनके मूल रूप (root form) में बदलना
- **Syntax & Semantics Analysis** – भाषा की संरचना और अर्थ का विश्लेषण

उदाहरण (Example)

यदि उपयोगकर्ता खोजता है—
"भारत में डिजिटल पुस्तकालय"
तो NLP आधारित IR प्रणाली देश (India), विषय (Digital Libraries) और संदर्भ (Context) को समझकर खोज परिणाम प्रस्तुत करती है।

12.8 Semantic Search एवं Context-based Retrieval:

Semantic Search ऐसी खोज तकनीक है जो केवल शब्दों (keywords) पर नहीं, बल्कि उनके **अर्थ (meaning)**, **संबंध (relationships)** और **संदर्भ (context)** पर आधारित होती है। यह तकनीक खोज प्रक्रिया को अधिक बुद्धिमान और उपयोगकर्ता-केंद्रित बनाती है।

Semantic Search की प्रमुख विशेषताएँ

- **Ontologies का प्रयोग** – ज्ञान के आपसी संबंधों को परिभाषित करना
- **Concept-based Retrieval** – अवधारणा के आधार पर खोज

- **Ambiguity Resolution** – बहुअर्थी शब्दों की अस्पष्टता का समाधान

उदाहरण (Example)

यदि उपयोगकर्ता खोजता है—
"Java"
तो Semantic Search प्रणाली संदर्भ के आधार पर तय करती है कि उपयोगकर्ता **Programming Language** के बारे में खोज रहा है या **Coffee** के बारे में।

12.9 Recommender Systems एवं Personalization

Recommender Systems ऐसी प्रणालियाँ हैं जो उपयोगकर्ता की **रुचि, पूर्व खोज व्यवहार और प्राथमिकताओं** के आधार पर प्रासंगिक सूचना या संसाधनों की सिफारिश (recommendation) करती हैं। ये प्रणालियाँ सूचना पुनर्प्राप्ति को **personalized** बनाती हैं।

Recommender Systems के प्रमुख प्रकार

- **Content-based Recommendation** – उपयोगकर्ता की रुचि के अनुसार
- **Collaborative Filtering** – समान रुचि वाले अन्य उपयोगकर्ताओं के आधार पर
- **Hybrid Systems** – Content-based और Collaborative दोनों का संयोजन

उदाहरण (Example)

Google Scholar उपयोगकर्ता की खोज और पढ़े गए लेखों के आधार पर **संबंधित शोध लेखों (related research papers)** की सिफारिश करता है।

12.10 Intelligent Search Engines

आधुनिक **Search Engines** पूर्णतः **बुद्धिमान सूचना पुनर्प्राप्ति (Intelligent Information Retrieval)** प्रणालियों पर आधारित हैं। ये search engines उपयोगकर्ता की खोज मंशा (user intent) को समझकर **सटीक, प्रासंगिक और त्वरित परिणाम** प्रदान करते हैं।

प्रमुख Intelligent Search Engines

- **Google**
- **Bing**
- **Semantic Scholar**

इन search engines में निम्नलिखित तकनीकों का उपयोग किया जाता है—

- **AI-based Ranking Algorithms**
- **User Intent Analysis**
- **Real-time Learning एवं Adaptation**

12.11 बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियों के लाभ

बुद्धिमान IR प्रणालियाँ पारंपरिक प्रणालियों की तुलना में अनेक लाभ प्रदान करती हैं—

- **उच्च प्रासंगिकता (High Relevance)**
- **समय की बचत (Time Efficiency)**
- **व्यक्तिगत खोज अनुभव (Personalized Search Experience)**
- **बहुभाषी समर्थन (Multilingual Support)**
- **जटिल एवं संदर्भ-आधारित प्रश्नों का समाधान**

12.12 सीमाएँ एवं चुनौतियाँ

यद्यपि बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियाँ अत्यंत प्रभावी हैं, फिर भी इनके समक्ष कुछ महत्वपूर्ण चुनौतियाँ विद्यमान हैं—

- **उच्च लागत (High Cost of Development & Maintenance)**
- **Privacy एवं Ethical Issues**
- **Algorithmic Bias (एल्गोरिदमिक पक्षपात)**
- **Multilingual एवं Cultural Complexity**
- **Skilled Manpower की आवश्यकता**

12.13 व्यावहारिक उदाहरण: Intelligent Information Retrieval System

Library Discovery Tools बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियों का व्यावहारिक उदाहरण हैं, जिनका उपयोग आधुनिक पुस्तकालयों में व्यापक रूप से किया जा रहा है।

उदाहरण: *EBSCO Discovery Service*

इस प्रकार के Intelligent IR systems में निम्नलिखित विशेषताएँ होती हैं—

- **Natural Language Query Processing** – उपयोगकर्ता की सामान्य भाषा में दी गई खोज को समझना

- **Faceted Search** – विषय, लेखक, वर्ष आदि के आधार पर परिणामों को फ़िल्टर करना
- **Personalized Results** – उपयोगकर्ता की रुचि एवं पूर्व खोज व्यवहार के अनुसार परिणाम प्रदान करना

इस प्रकार, Library Discovery Tools उपयोगकर्ताओं को **तेज़, सटीक और उपयोगकर्ता-केंद्रित खोज अनुभव** प्रदान करते हैं।

12.14 भविष्य की प्रवृत्तियाँ (Future Trends)

बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियाँ निरंतर विकसित हो रही हैं। भविष्य में इस क्षेत्र में निम्नलिखित प्रवृत्तियाँ विशेष रूप से महत्वपूर्ण होंगी—

- **AI-powered Chatbots** के माध्यम से संवादात्मक खोज
- **Voice-based Search** द्वारा ध्वनि आधारित सूचना पुनर्प्राप्ति
- **Emotion-aware Retrieval** के माध्यम से उपयोगकर्ता की भावनात्मक स्थिति के अनुरूप परिणाम
- **Large Language Models (LLMs)** के साथ एकीकरण
- **Fully Adaptive IR Systems**, जो समय के साथ स्वयं को अनुकूलित कर सकें

12.15 सारांश (Summary)

बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियाँ (Intelligent Information Retrieval Systems) आधुनिक सूचना समाज की एक **महत्वपूर्ण आधारशिला (backbone)** बन चुकी हैं। ये प्रणालियाँ पारंपरिक सूचना पुनर्प्राप्ति (Traditional IR) की सीमाओं को पार करते हुए केवल **keyword-based search** के स्थान पर **meaning, context और user experience** को केंद्र में रखती हैं, जिससे खोज परिणाम अधिक **relevant, accurate और useful** बनते हैं।

Library एवं Information Science के क्षेत्र में, विशेषकर **digital libraries, online databases और research information systems** में, इन प्रणालियों का महत्व निरंतर बढ़ रहा है और ये आधुनिक पुस्तकालय सेवाओं को अधिक **efficient, user-centred और intelligent** बनाने में महत्वपूर्ण भूमिका निभा रही हैं।

12.16 शब्दावली (Glossary)

(Key Terms Related to Intelligent Information Retrieval Systems)

- **Information Retrieval (IR):** विशाल सूचना संग्रह से उपयोगकर्ता की आवश्यकता के अनुरूप प्रासंगिक सूचना को खोजने एवं प्राप्त करने की प्रक्रिया।
- **Artificial Intelligence (AI) :** कंप्यूटर प्रणालियों को मानव-सदृश बुद्धिमत्ता प्रदान करने की तकनीक, जिससे वे सीखने, निर्णय लेने और समस्या समाधान में सक्षम बनती हैं।
- **Machine Learning (ML):** Artificial Intelligence की वह शाखा जिसमें प्रणाली अनुभव एवं डेटा के आधार पर स्वयं सीखती और अपनी कार्यक्षमता में सुधार करती है।
- **Natural Language Processing (NLP):** वह तकनीक जो मशीन को मानव भाषा (text या speech) को समझने, विश्लेषण करने और संसाधित करने में सक्षम बनाती है।
- **Semantic Search:** ऐसी खोज प्रक्रिया जो केवल keywords पर नहीं, बल्कि शब्दों के अर्थ (meaning), संबंध (relationship) और संदर्भ (context) पर आधारित होती है।
- **Context-based Retrieval:** उपयोगकर्ता के संदर्भ, स्थान, समय एवं खोज उद्देश्य के आधार पर सूचना पुनर्प्राप्ति की प्रक्रिया।
- **Ontology:** किसी विषय क्षेत्र में अवधारणाओं (concepts) तथा उनके आपसी संबंधों का संरचित निरूपण।
- **Recommender System:** उपयोगकर्ता की रुचि, पूर्व व्यवहार एवं प्राथमिकताओं के आधार पर सूचना या संसाधनों की सिफारिश करने वाली प्रणाली।
- **Personalization:** उपयोगकर्ता-विशिष्ट आवश्यकताओं एवं रुचियों के अनुसार खोज परिणामों को अनुकूलित करने की प्रक्रिया।
- **User Behaviour Analysis:** उपयोगकर्ता की खोज आदतों, पसंद और प्रतिक्रिया का अध्ययन, जिससे खोज परिणामों को बेहतर बनाया जा सके।
- **Intelligent Search Engine:** ऐसा search engine जो AI, ML और NLP का प्रयोग करके अधिक सटीक, प्रासंगिक और personalized परिणाम प्रदान करता है।

12.17 निबंधात्मक प्रश्न (Essay Type Questions with Answers)

प्रश्न 1. बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली की अवधारणा स्पष्ट कीजिए।

उत्तर: बुद्धिमान सूचना पुनर्प्राप्ति प्रणाली (Intelligent Information Retrieval System) ऐसी प्रणाली है जो केवल कीवर्ड के आधार पर सूचना खोजने तक सीमित नहीं रहती, बल्कि **अर्थ (meaning), संदर्भ (context), उपयोगकर्ता की सूचना आवश्यकता (information**

need) **तथा उपयोगकर्ता व्यवहार (user behaviour)** को ध्यान में रखकर खोज परिणाम प्रदान करती है।

इन प्रणालियों में **Artificial Intelligence, Machine Learning, Natural Language Processing और Semantic Technologies** का प्रयोग किया जाता है, जिससे प्रणाली उपयोगकर्ता की मंशा को समझने में सक्षम होती है। परिणामस्वरूप, खोज परिणाम अधिक **प्रासंगिक, सटीक एवं व्यक्तिगत (personalized)** बनते हैं। आधुनिक search engines, library discovery tools और recommender systems बुद्धिमान सूचना पुनर्प्राप्ति के प्रमुख उदाहरण हैं।

प्रश्न 2. पारंपरिक IR और बुद्धिमान IR में अंतर बताइए।

उत्तर: पारंपरिक सूचना पुनर्प्राप्ति प्रणालियाँ मुख्यतः **keyword matching** पर आधारित होती हैं, जहाँ खोज परिणाम सभी उपयोगकर्ताओं के लिए लगभग समान होते हैं। इनमें उपयोगकर्ता के संदर्भ या व्यवहार को महत्व नहीं दिया जाता।

इसके विपरीत, बुद्धिमान सूचना पुनर्प्राप्ति प्रणालियाँ **अर्थ, संदर्भ और उपयोगकर्ता प्रोफ़ाइल** को ध्यान में रखती हैं। ये प्रणालियाँ समय के साथ सीखती हैं और खोज परिणामों को उपयोगकर्ता की रुचि के अनुसार अनुकूलित करती हैं। इसी कारण बुद्धिमान IR प्रणालियाँ पारंपरिक प्रणालियों की तुलना में अधिक प्रभावी मानी जाती हैं।

प्रश्न 3. सूचना पुनर्प्राप्ति में AI और ML की भूमिका पर चर्चा कीजिए।

उत्तर: Artificial Intelligence (AI) और Machine Learning (ML) ने सूचना पुनर्प्राप्ति प्रणालियों को अधिक उन्नत और बुद्धिमान बनाया है। AI खोज प्रश्नों के अर्थ और संदर्भ को समझने में सहायता करता है, जबकि ML उपयोगकर्ता के पूर्व खोज व्यवहार और प्रतिक्रिया से सीखकर भविष्य के खोज परिणामों को बेहतर बनाता है।

AI के माध्यम से **intelligent ranking, reasoning और user intent analysis** संभव हो पाता है, वहीं ML द्वारा **learning to rank, query expansion और document classification** जैसे कार्य किए जाते हैं। इन तकनीकों के कारण खोज प्रणाली समय के साथ अधिक सटीक और उपयोगकर्ता-केंद्रित बनती जाती है।

12.18 लघु प्रश्न (Short Answer Questions with Answers)

1. Semantic Search क्या है?

Semantic Search ऐसी खोज तकनीक है जो शब्दों के स्थान पर उनके **अर्थ और संदर्भ** के आधार पर खोज परिणाम प्रदान करती है।

2. Recommender System का अर्थ बताइए।

Recommender System वह प्रणाली है जो उपयोगकर्ता की रुचि और व्यवहार के आधार पर सूचना या संसाधनों की सिफारिश करती है।

12.18 (D) बहुविकल्पीय प्रश्न (MCQs)

1. Intelligent IR system का मुख्य आधार क्या है?
   a) Storage
   b) Keyword matching
   c) Meaning & context ✔
   d) Hardware

2. Semantic Search का उद्देश्य क्या है?
   a) Faster storage
   b) Meaning-based retrieval ✔
   c) Data compression
   d) Index removal

3. NLP का प्रयोग IR में किसलिए होता है?
   a) Hardware control
   b) Language understanding ✔
   c) Memory management
   d) File deletion

4. Learning to Rank किससे संबंधित है?
   a) NLP
   b) Ontology
   c) Machine Learning ✔
   d) Database design

5. Recommender Systems का मुख्य उद्देश्य है—
   a) Data storage
   b) User personalization ✔
   c) Indexing
   d) Classification

6. User profiling का प्रयोग किसमें होता है?
   a) Traditional IR
   b) Intelligent IR ✔
   c) Manual indexing
   d) OPAC only

7. Ontology का प्रयोग मुख्यतः किसमें होता है?
   a) Semantic Search ✔
   b) Spam filtering
   c) Hardware design
   d) Backup

8. Intelligent Search Engine का उदाहरण है—
   a) Notepad
   b) Google ✔
   c) MS Paint
   d) Calculator

9. Algorithmic bias किससे संबंधित है?
   a) Hardware failure
   b) Ethical issues ✔
   c) Storage error
   d) Network speed

10. Voice-based search किस future trend का उदाहरण है?
   a) Manual IR
   b) Traditional IR
   c) Intelligent IR ✔
   d) Offline IR

12.20 संदर्भ ग्रंथ सूची (References)

1. Salton, G. Introduction to Modern Information Retrieval.
2. Manning, C. D., Raghavan, P., & Schütze, H. Introduction to Information Retrieval.
3. Baeza-Yates, R., & Ribeiro-Neto, B. Modern Information Retrieval.
4. Chowdhury, G. G. Information Retrieval Systems.
5. Hearst, M. Search User Interfaces.
6. Croft, W. B., Metzler, D., & Strohman, T. Search Engines: Information Retrieval in Practice.
7. Jurafsky, D., & Martin, J. Speech and Language Processing.
8. Russell, S., & Norvig, P. Artificial Intelligence: A Modern Approach.
9. Ricci, F., Rokach, L., & Shapira, B. Recommender Systems Handbook.
10. Feldman, R., & Sanger, J. The Text Mining Handbook.
11. रंगनाथन, एस. आर. – सूचना संगठन एवं पुनर्प्राप्ति
12. गुप्ता, एस. के. – पुस्तकालय स्वचालन एवं सूचना प्रौद्योगिकी
13. मिश्र, आर. एन. – डिजिटल पुस्तकालय
14. वर्मा, आर. के. – सूचना विज्ञान के सिद्धांत

15. शर्मा, बी. एल. – सूचना पुनर्प्राप्ति प्रणालियाँ
16. सिंह, पी. – आधुनिक पुस्तकालय सेवाएँ
17. पाण्डेय, एस. – सूचना एवं संचार प्रौद्योगिकी
18. तिवारी, आर. – डिजिटल सूचना प्रणाली
19. यादव, के. – पुस्तकालय एवं सूचना विज्ञान
20. हिंदी अकादमिक प्रकाशन – डिजिटल पुस्तकालय एवं सूचना सेवा

इकाई–13 : सूचना पुनर्प्राप्ति : अवधारणाएँ एवं तकनीकें (Information Retrieval – Concepts and Techniques)

इकाई की रूपरेखा

13.1 प्रस्तावना
13.2 उद्देश्य
13.3 सूचना पुनर्प्राप्ति (Information Retrieval) : अर्थ एवं परिभाषा
13.4 सूचना पुनर्प्राप्ति प्रणाली के घटक
13.4.1 सूचना संसाधन (Information Sources)
13.4.2 उपयोगकर्ता (Users)
13.4.3 खोज इंटरफेस (Search Interface)
13.4.4 पुनर्प्राप्ति तंत्र (Retrieval Mechanism)

13.5 सूचना पुनर्प्राप्ति प्रक्रिया (IR Process)
13.6 सूचना आवश्यकता (Information Need) एवं प्रासंगिकता (Relevance)
13.7 अनुक्रमण (Indexing) : अवधारणा एवं तकनीकें
13.7.1 मैनुअल अनुक्रमण
13.7.2 स्वचालित अनुक्रमण

13.8 खोज तकनीकें (Search Techniques)
13.8.1 की-वर्ड खोज
13.8.2 विषय आधारित खोज
13.8.3 बूलियन खोज तकनीक

13.9 पुनर्प्राप्ति की मापन विधियाँ (Evaluation of IR Systems)
13.9.1 Precision
13.9.2 Recall
13.9.3 F-measure

13.10 सूचना पुनर्प्राप्ति में स्वचालन की भूमिका
13.11 डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति
13.12 सूचना पुनर्प्राप्ति की सीमाएँ एवं चुनौतियाँ
13.13 व्यावहारिक उदाहरण: OPAC एवं डेटाबेस खोज

13.14 नवीन प्रवृत्तियाँ: AI आधारित IR सिस्टम
13.15 सारांश
13.16 शब्दावली
13.17 निबंधात्मक प्रश्न
13.18 लघु प्रश्न एवं MCQ
13.19 अभ्यास प्रश्नों के उत्तर
13.20 संदर्भ ग्रंथ सूची

13.1 प्रस्तावना (Introduction) :

सूचना युग (Information Age) में ज्ञान का सृजन एवं संप्रेषण अत्यंत तीव्र गति से हो रहा है। पुस्तकें, शोध-पत्र, ई-जर्नल, डेटाबेस, वेब संसाधन एवं डिजिटल रिपॉजिटरी के रूप में उपलब्ध इस विशाल सूचना-संसार में से **उपयोगकर्ता की आवश्यकता के अनुरूप प्रासंगिक सूचना को खोजकर प्रस्तुत करना** ही **सूचना पुनर्प्राप्ति (Information Retrieval – IR)** का मूल उद्देश्य है।

पारंपरिक पुस्तकालय व्यवस्था में सूचना की खोज कार्ड कैटलॉग, वर्गीकरण योजनाओं एवं विषय सूचियों के माध्यम से की जाती थी। किंतु आधुनिक डिजिटल परिवेश में यह कार्य **OPAC, ऑनलाइन डेटाबेस, सर्च इंजन तथा डिजिटल लाइब्रेरी प्रणालियों** द्वारा अधिक प्रभावी, तीव्र और सटीक रूप में संपन्न किया जाता है।

यह इकाई सूचना पुनर्प्राप्ति की **मूल अवधारणाओं, प्रक्रियाओं, तकनीकों, मूल्यांकन विधियों तथा नवीन प्रवृत्तियों** का सरल, व्यावहारिक एवं शिक्षार्थी-केंद्रित विवेचन प्रस्तुत करती है, जिससे शिक्षार्थी सूचना खोज की आधुनिक प्रणाली को समग्र रूप से समझ सकें।

13.2 उद्देश्य (Objectives) :

इस इकाई के अध्ययन के उपरांत शिक्षार्थी—

- सूचना पुनर्प्राप्ति की अवधारणा को स्पष्ट रूप से समझ सकेंगे
- सूचना पुनर्प्राप्ति प्रणाली के प्रमुख घटकों की पहचान कर सकेंगे
- सूचना आवश्यकता एवं प्रासंगिकता के संबंध को विश्लेषित कर सकेंगे
- मैनुअल एवं स्वचालित अनुक्रमण के बीच अंतर स्पष्ट कर सकेंगे
- विभिन्न खोज तकनीकों (Keyword, Boolean आदि) का प्रभावी उपयोग कर सकेंगे
- Precision, Recall तथा F-measure को उदाहरण सहित समझ सकेंगे

- डिजिटल पुस्तकालयों एवं AI आधारित सूचना पुनर्प्राप्ति प्रणालियों की भूमिका का मूल्यांकन कर सकेंगे

13.3 सूचना पुनर्प्राप्ति : अर्थ एवं परिभाषा (Information Retrieval: Meaning and Definition):

सूचना पुनर्प्राप्ति (Information Retrieval – IR) वह प्रक्रिया है जिसके अंतर्गत संग्रहीत सूचना-संसाधनों (Information Resources) में से **उपयोगकर्ता की सूचना-आवश्यकता (Information Need)** के अनुरूप **सर्वाधिक प्रासंगिक सूचना (Relevant Information)** को खोजकर उपलब्ध कराया जाता है।

सरल शब्दों में—

"सूचना की विशाल भीड़ में से उपयोगी और आवश्यक सूचना को खोज निकालने की प्रक्रिया को सूचना पुनर्प्राप्ति कहा जाता है।"

(English Definition)

According to Chowdhury (2004): Information Retrieval is the process of representing, storing, organizing and accessing information items in order to retrieve those items that satisfy a user's information need.

इस परिभाषा से स्पष्ट होता है कि सूचना पुनर्प्राप्ति केवल खोज (search) तक सीमित नहीं है, बल्कि इसमें **सूचना का संगठन (organization), अनुक्रमण (indexing) और विश्लेषण (analysis)** भी सम्मिलित होता है।

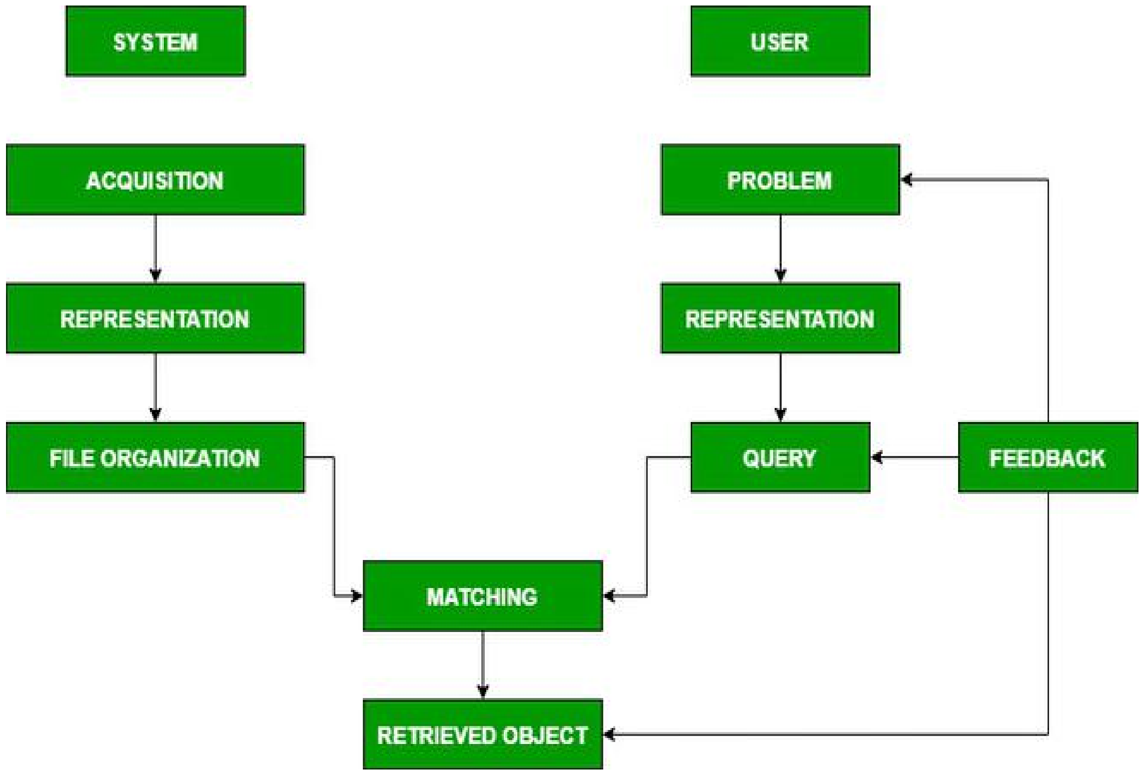
उदाहरण (Examples)

- किसी शोधार्थी द्वारा *"Artificial Intelligence in Libraries"* विषय पर शोध-लेख खोजने हेतु **online database** का उपयोग
- **OPAC (Online Public Access Catalogue)** में किसी पुस्तक को **शीर्षक (Title)** या **लेखक (Author)** के आधार पर खोज

13.4 सूचना पुनर्प्राप्ति प्रणाली के घटक (Components of Information Retrieval System):

सूचना पुनर्प्राप्ति प्रणाली (Information Retrieval System – IR System) एक सुव्यवस्थित एवं समन्वित तंत्र है, जो विभिन्न घटकों के पारस्परिक सहयोग से कार्य करता है। किसी भी IR

System की प्रभावशीलता उसके घटकों के संतुलित एवं संगठित संचालन पर निर्भर करती है। सामान्यतः एक सूचना पुनर्प्राप्ति प्रणाली **चार प्रमुख घटकों** पर आधारित होती है।



(Components of Information Retrieval System)

13.4.1 सूचना संसाधन (Information Sources)

सूचना संसाधन वे सभी माध्यम हैं जिनमें सूचना संग्रहीत होती है तथा जिनसे उपयोगकर्ता अपनी सूचना-आवश्यकता की पूर्ति करता है। ये संसाधन सूचना पुनर्प्राप्ति प्रणाली के **मूल आधार (core element)** होते हैं। प्रमुख सूचना संसाधनों में निम्नलिखित शामिल हैं—

- पुस्तकें (Books)
- जर्नल एवं शोध-पत्र (Journals & Research Articles)
- शोध-प्रबंध (Theses & Dissertations)
- ई-संसाधन (E-resources)
- मल्टीमीडिया संसाधन (Multimedia Resources)

सूचना संसाधनों की गुणवत्ता, संगठन एवं उपलब्धता सीधे-सीधे सूचना पुनर्प्राप्ति प्रणाली की कार्यक्षमता और परिणामों को प्रभावित करती है। बिना उपयुक्त सूचना संसाधनों के किसी भी IR System की पुनर्प्राप्ति प्रक्रिया प्रभावी नहीं हो सकती।

13.4.2 उपयोगकर्ता (Users)

उपयोगकर्ता (Users) वे व्यक्ति होते हैं जिनकी **सूचना आवश्यकता (Information Need)** को पूरा करने के लिए सूचना पुनर्प्राप्ति प्रणाली विकसित की जाती है। किसी भी IR System की सफलता इस बात पर निर्भर करती है कि वह उपयोगकर्ता की आवश्यकता को कितनी प्रभावी एवं सटीक रूप से पूरा कर पाता है।

प्रमुख उपयोगकर्ता वर्ग (Major User Groups):

- विद्यार्थी (Students)
- शोधार्थी (Researchers)
- शिक्षक (Teachers / Faculty Members)
- पुस्तकालय कर्मी (Library Professionals)

English Definition

Users are individuals who interact with an Information Retrieval System to satisfy their information needs by searching and retrieving relevant information.

13.4.3 खोज इंटरफेस (Search Interface)

खोज इंटरफेस (Search Interface) वह माध्यम है जिसके द्वारा उपयोगकर्ता सूचना पुनर्प्राप्ति प्रणाली के साथ संवाद (interaction) करता है। यह उपयोगकर्ता और प्रणाली के बीच **communication link** के रूप में कार्य करता है।

प्रमुख खोज इंटरफेस के उदाहरण:

- OPAC (Online Public Access Catalogue)
- डेटाबेस सर्च विंडो (Database Search Window)
- वेब सर्च इंजन (Web Search Engines)

एक अच्छा search interface **user-friendly, interactive और flexible** होना चाहिए ताकि उपयोगकर्ता प्रभावी ढंग से खोज कर सके।

English Definition:

Search Interface is the point of interaction between the user and the information retrieval system through which search queries are formulated and results are displayed.

13.4.4 पुनर्प्राप्ति तंत्र (Retrieval Mechanism)

पुनर्प्राप्ति तंत्र (Retrieval Mechanism) वह तकनीकी प्रक्रिया है जिसके माध्यम से प्रणाली उपयोगकर्ता की query को संसाधित कर **प्रासंगिक सूचना (Relevant Information)** को खोजकर प्रस्तुत करती है।

पुनर्प्राप्ति तंत्र के प्रमुख घटक:

- अनुक्रमण (Indexing)
- खोज एल्गोरिद्म (Search Algorithms)
- रैंकिंग प्रणाली (Ranking System)

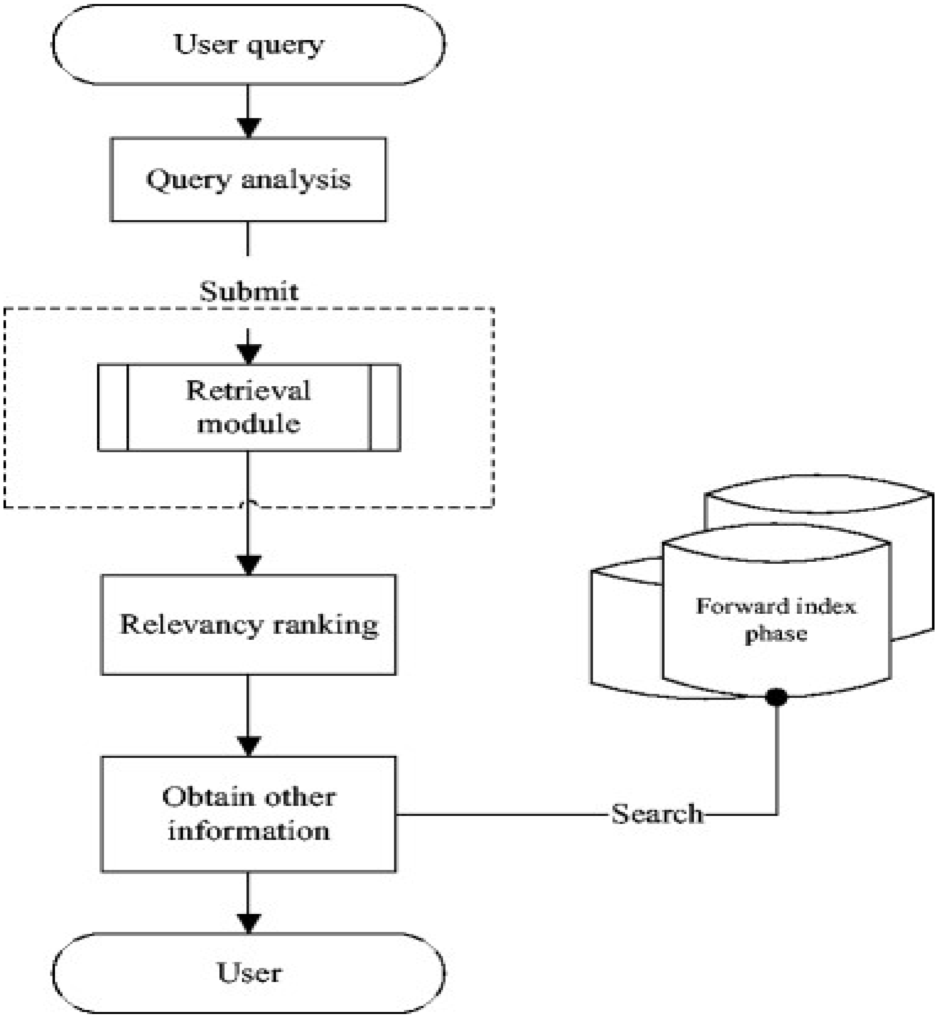
यह तंत्र यह निर्धारित करता है कि **कौन-सी सूचना पहले प्रदर्शित होगी और क्यों।**

English Definition

Retrieval Mechanism refers to the set of techniques and algorithms used by an information retrieval system to match user queries with stored information and rank the retrieved results.

13.5 सूचना पुनर्प्राप्ति प्रक्रिया (IR Process):

सूचना पुनर्प्राप्ति एक **क्रमिक (sequential)** और **तार्किक (logical)** प्रक्रिया है, जिसके माध्यम से उपयोगकर्ता की सूचना-आवश्यकता को पहचानकर उसे **प्रासंगिक सूचना (relevant information)** में परिवर्तित किया जाता है। यह प्रक्रिया उपयोगकर्ता और सूचना संसाधनों के बीच एक प्रभावी सेतु का कार्य करती है।



(Information Retrieval Process)

सूचना पुनर्प्राप्ति प्रक्रिया सामान्यतः निम्नलिखित चरणों में पूर्ण होती है—

1. **सूचना आवश्यकता की पहचान (Identification of Information Need)** इस चरण में उपयोगकर्ता अपनी समस्या या जिज्ञासा के आधार पर यह निर्धारित करता है कि उसे किस प्रकार की सूचना की आवश्यकता है। यह आवश्यकता स्पष्ट भी हो सकती है और अस्पष्ट भी।
2. **खोज प्रश्न का निर्माण (Formulation of Search Query)** सूचना आवश्यकता को उपयुक्त **keywords, subject terms** या **Boolean operators** के माध्यम से खोज-प्रश्न (query) के रूप में व्यक्त किया जाता है।

3. **अनुक्रमण एवं मिलान (Indexing and Matching):** प्रणाली दस्तावेज़ों के अनुक्रमण (indexes) का उपयोग कर खोज-प्रश्न और संग्रहित सूचना के बीच मिलान करती है।
4. **प्रासंगिक परिणामों की प्राप्ति (Retrieval of Relevant Results):** मिलान के आधार पर प्रणाली उपयोगकर्ता को संभावित रूप से प्रासंगिक दस्तावेज़ प्रस्तुत करती है।
5. **परिणामों का मूल्यांकन (Evaluation of Results):** उपयोगकर्ता प्राप्त परिणामों की उपयोगिता और प्रासंगिकता का मूल्यांकन करता है तथा आवश्यकता होने पर खोज-प्रश्न में संशोधन करता है।

इस प्रकार सूचना पुनर्प्राप्ति प्रक्रिया **iterative** होती है, अर्थात आवश्यकता पड़ने पर इसे दोहराया भी जा सकता है।

13.6 सूचना आवश्यकता एवं प्रासंगिकता (Information Need and Relevance):

सूचना आवश्यकता (Information Need)

सूचना आवश्यकता वह **बौद्धिक आवश्यकता (intellectual need)** है, जिसे उपयोगकर्ता किसी समस्या के समाधान, निर्णय-निर्माण या ज्ञानवृद्धि के लिए अनुभव करता है। यह आवश्यकता प्रारंभ में अस्पष्ट हो सकती है, जो खोज प्रक्रिया के दौरान धीरे-धीरे स्पष्ट होती जाती है।

प्रासंगिकता (Relevance)

प्रासंगिकता (Relevance) उस स्तर को दर्शाती है जिस तक कोई सूचना-संसाधन उपयोगकर्ता की सूचना-आवश्यकता को पूरा करता है। खोज परिणाम तभी उपयोगी माने जाते हैं जब वे उपयोगकर्ता की आवश्यकता से **सार्थक रूप से संबद्ध** हों।

उदाहरण:
यदि किसी उपयोगकर्ता को *"डिजिटल लाइब्रेरी सॉफ्टवेयर"* से संबंधित जानकारी चाहिए, तो **KOHA**, **DSpace** अथवा अन्य ओपन-सोर्स डिजिटल लाइब्रेरी प्लेटफॉर्म पर आधारित दस्तावेज़ प्रासंगिक माने जाएँगे, जबकि असंबंधित सॉफ्टवेयर पर सामग्री अप्रासंगिक होगी।

13.7 अनुक्रमण : अवधारणा एवं तकनीकें (Indexing: Concept and Techniques):

अनुक्रमण (Indexing) वह प्रक्रिया है जिसके अंतर्गत किसी दस्तावेज़ की विषयवस्तु का विश्लेषण कर उसे **खोज-योग्य (searchable)** रूप में अभिव्यक्त किया जाता है। प्रभावी अनुक्रमण सूचना पुनर्प्राप्ति प्रणाली की **आधारशिला (foundation)** मानी जाती है, क्योंकि खोज की सफलता काफी हद तक अनुक्रमण की गुणवत्ता पर निर्भर करती है।

13.7.1 मैनुअल अनुक्रमण (Manual Indexing)

मैनुअल अनुक्रमण मानव द्वारा किया जाता है, जिसमें विषय विशेषज्ञ दस्तावेज़ की विषयवस्तु को समझकर उपयुक्त शब्दों का चयन करता है। यह विधि अधिक **सटीक (accurate)** होती है, किंतु **समयसाध्य (time-consuming)** होती है।

13.7.2 स्वचालित अनुक्रमण (Automatic Indexing)

स्वचालित अनुक्रमण कंप्यूटर-आधारित तकनीकों द्वारा किया जाता है, जिसमें एल्गोरिद्म शब्द आवृत्ति, सांख्यिकीय विश्लेषण आदि के आधार पर अनुक्रमण करता है। यह विधि **तेज (fast)** होती है और बड़े संग्रह के लिए उपयुक्त है, परंतु कभी-कभी **कम सटीक (less accurate)** हो सकती है।

13.8 खोज तकनीकें (Search Techniques)

खोज तकनीकें (Search Techniques) वे विधियाँ हैं जिनके माध्यम से उपयोगकर्ता सूचना पुनर्प्राप्ति प्रणाली में **प्रभावी (effective)** एवं **सटीक (accurate)** खोज कर सकता है। उपयुक्त खोज तकनीक का चयन खोज परिणामों की गुणवत्ता को प्रत्यक्ष रूप से प्रभावित करता है।

13.8.1 की-वर्ड खोज (Keyword Search)

की-वर्ड खोज वह तकनीक है जिसमें उपयोगकर्ता **सीधे शब्दों (keywords)** के आधार पर खोज करता है। यह सबसे सरल एवं सामान्य खोज विधि है।

विशेषताएँ (Features):

- सरल और तेज (simple and fast)
- नियंत्रित शब्दावली की आवश्यकता नहीं
- परिणाम अधिक हो सकते हैं, परंतु कभी-कभी अप्रासंगिक भी

उदाहरण:
"Digital Library" शब्द से डेटाबेस में खोज करना।

13.8.2 विषय आधारित खोज (Subject-based Search)

विषय आधारित खोज में नियंत्रित शब्दावली **(controlled vocabulary)** या थिसॉरस **(thesaurus)** का उपयोग किया जाता है। यह खोज अधिक सटीक **(precise)** मानी जाती है।

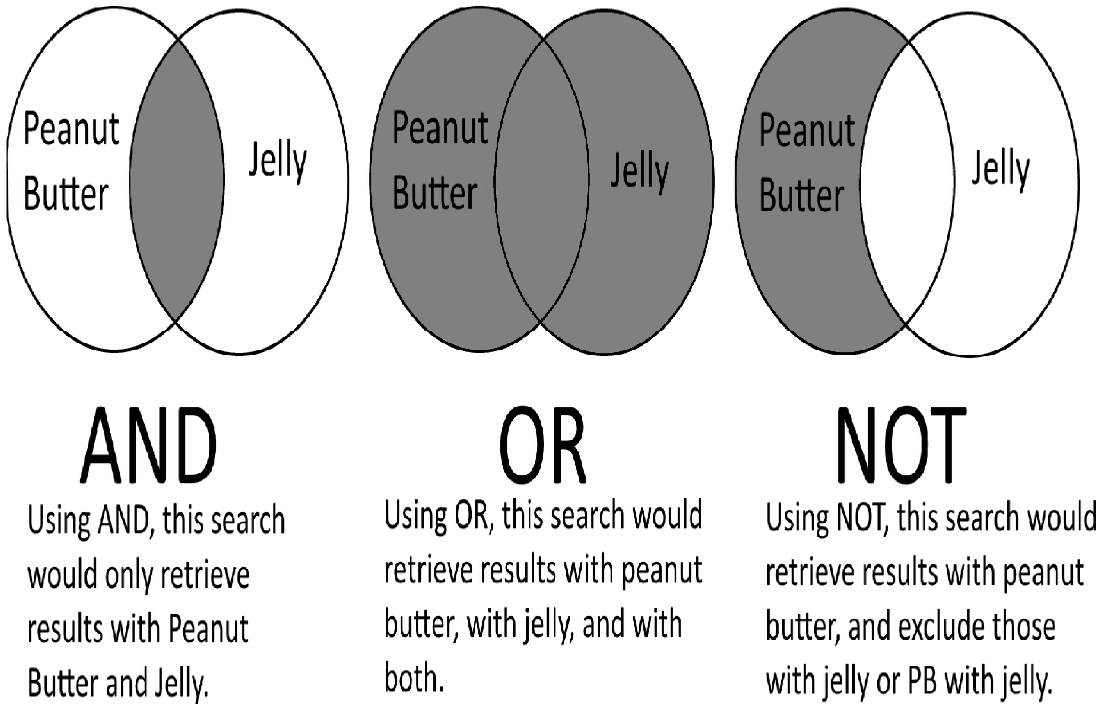
विशेषताएँ:

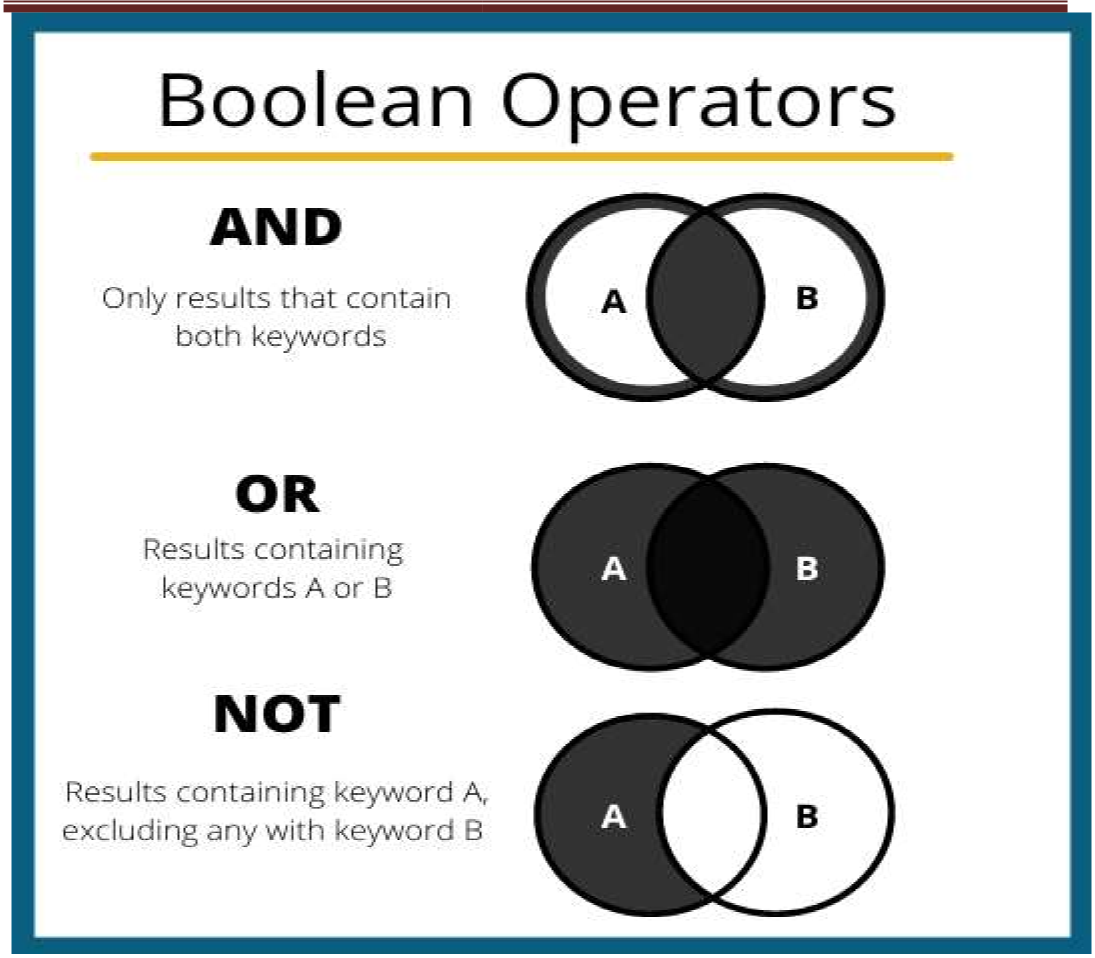
- मानकीकृत विषय शब्दों का प्रयोग
- अप्रासंगिक परिणाम कम
- अकादमिक डेटाबेस में अधिक उपयोगी

उदाहरण:
Library of Congress Subject Headings (LCSH) या Thesaurus का प्रयोग।

13.8.3 बूलियन खोज तकनीक (Boolean Search Technique)

बूलियन खोज तकनीक में **Boolean operators** का उपयोग कर खोज को नियंत्रित एवं परिष्कृत किया जाता है।





प्रमुख Boolean Operators:

- **AND** – खोज को **संकुचित (narrow)** करता है
- **OR** – खोज को **विस्तृत (broaden)** करता है
- **NOT** – **अनावश्यक परिणामों (irrelevant results)** को हटाता है

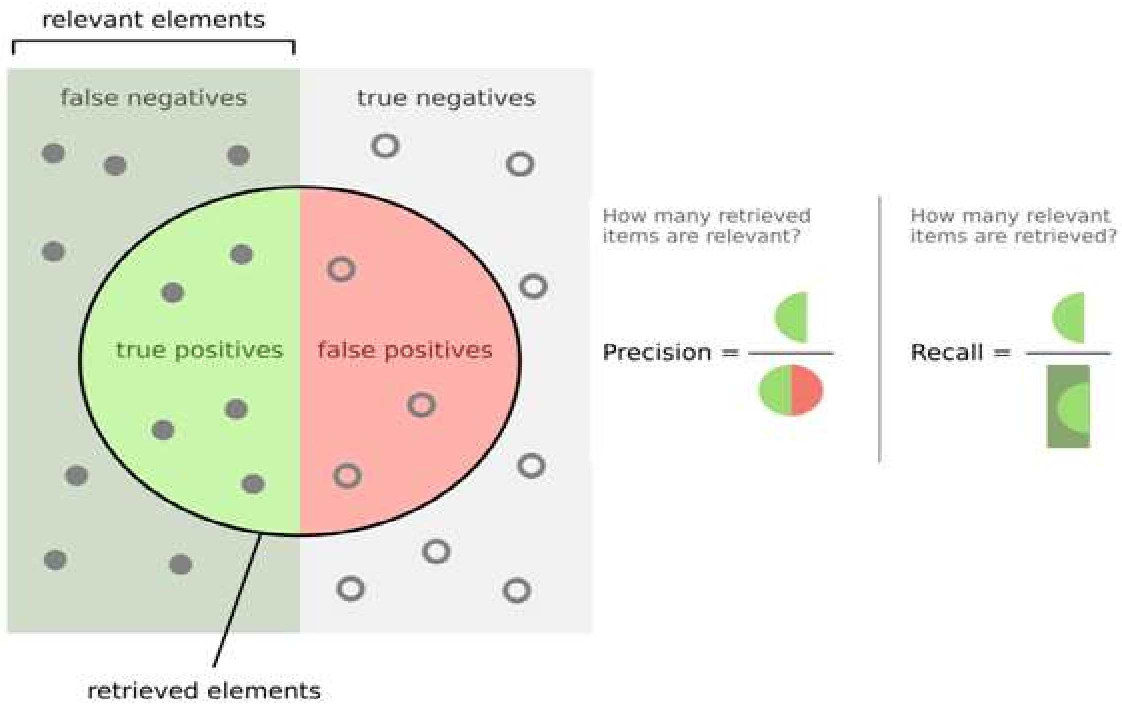
उदाहरण:

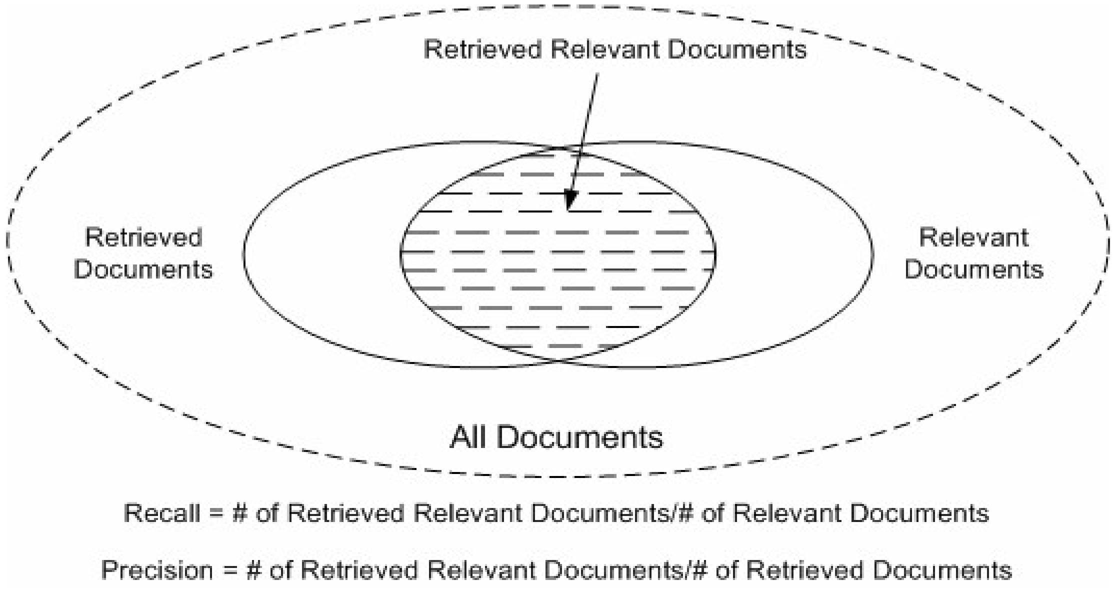
- Library AND Automation
- Digital OR Electronic
- Library NOT School

- AND – खोज को संकुचित करता है
- OR – खोज को विस्तृत करता है
- NOT – अनावश्यक परिणाम हटाता है

13.9 पुनर्प्राप्ति की मापन विधियाँ (Evaluation Measures of IR Systems)

सूचना पुनर्प्राप्ति प्रणाली की प्रभावशीलता को मापने के लिए कुछ मानक मापन विधियाँ **(standard evaluation measures)** प्रयुक्त की जाती हैं। इनमें प्रमुख रूप से **Precision, Recall** और **F-measure** शामिल हैं।





(i) Precision

Precision वह अनुपात है जो यह दर्शाता है कि **प्राप्त (retrieved) दस्तावेज़ों में से कितने वास्तव में प्रासंगिक (relevant)** हैं।

English Definition:

Precision is the proportion of retrieved documents that are relevant.

सूत्र (Formula):

Precision = Relevant Retrieved Documents / Total Retrieved Documents

(ii) Recall

Recall वह अनुपात है जो यह बताता है कि **कुल उपलब्ध प्रासंगिक दस्तावेज़ों में से कितने पुनर्प्राप्त किए गए** हैं।

English Definition:

Recall is the proportion of relevant documents that are successfully retrieved.

सूत्र (Formula):

Recall = Relevant Retrieved Documents / Total Relevant Documents in the Collection

(iii) F-measure (F-score)

F-measure वह माप है जो **Precision और Recall के बीच संतुलन (balance)** को दर्शाता है। यह तब उपयोगी होता है जब दोनों को समान महत्व देना हो।

English Definition: ‘***F-measure is the harmonic mean of Precision and Recall.***’

13.10 सूचना पुनर्प्राप्ति में स्वचालन की भूमिका (Role of Automation in Information Retrieval)

आधुनिक सूचना पुनर्प्राप्ति प्रणालियों में **स्वचालन (automation)** की भूमिका अत्यंत महत्वपूर्ण है। कंप्यूटर आधारित तकनीकों के माध्यम से खोज प्रक्रिया अधिक **तेज, सटीक और उपयोगकर्ता-अनुकूल** बन गई है।

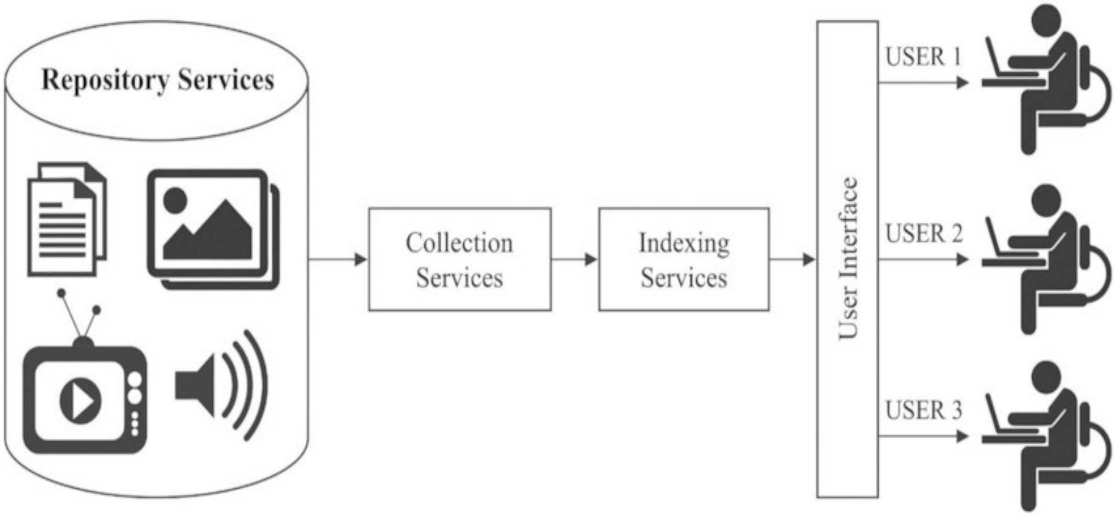
सूचना पुनर्प्राप्ति में स्वचालन के प्रमुख लाभ:

- **समय की बचत (Time efficiency)**
- **बड़े डेटाबेस में प्रभावी खोज (Effective searching in large databases)**
- **उपयोगकर्ता-अनुकूल इंटरफेस (User-friendly interfaces)**
- **तेज एवं स्वचालित अनुक्रमण (Automatic indexing and retrieval)**

स्वचालन के कारण डिजिटल पुस्तकालय, ऑनलाइन डेटाबेस और सर्च इंजन अधिक प्रभावी रूप से कार्य कर पा रहे हैं।

3.11 डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति (Information Retrieval in Digital Libraries)

डिजिटल पुस्तकालयों (Digital Libraries) में सूचना पुनर्प्राप्ति (Information Retrieval) की भूमिका अत्यंत महत्वपूर्ण है, क्योंकि डिजिटल वातावरण में सूचना की मात्रा, स्वरूप और उपयोगकर्ता अपेक्षाएँ निरंतर बढ़ रही हैं। प्रभावी IR प्रणाली उपयोगकर्ताओं को सटीक **(accurate)** एवं त्वरित **(efficient)** सूचना उपलब्ध कराने में सहायक होती है।



चित्र: डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति की रूपरेखा

डिजिटल पुस्तकालयों में IR के महत्व के प्रमुख कारण:

- **संसाधन विशाल होते हैं (Large and diverse collections)**
- **उपयोगकर्ता विविध होते हैं (Diverse user groups)**

- **खोज बहुआयामी होती है (Multidimensional searching)**
- **Metadata, indexing और search algorithms का व्यापक उपयोग**

डिजिटल पुस्तकालयों में OPAC, institutional repositories, e-journals और databases के माध्यम से उन्नत खोज सुविधाएँ उपलब्ध कराई जाती हैं।

13.12 सीमाएँ एवं चुनौतियाँ **(Limitations and Challenges of Information Retrieval):** **सूचना पुनर्प्राप्ति प्रणालियाँ अत्यधिक उपयोगी होने के बावजूद** कुछ सीमाओं एवं चुनौतियों (limitations and challenges) **से ग्रस्त होती हैं, जो खोज की गुणवत्ता को प्रभावित कर सकती हैं।**

प्रमुख सीमाएँ एवं चुनौतियाँ:

- **भाषा की अस्पष्टता (Language ambiguity)**
- **सूचना अधिभार (Information overload)**
- **प्रासंगिकता निर्धारण की समस्या (Difficulty in relevance judgment)**
- **उपयोगकर्ता की अस्पष्ट सूचना आवश्यकता (Unclear information need)**

इन चुनौतियों के कारण कभी-कभी उपयोगकर्ता को अपेक्षित सूचना प्राप्त करने में कठिनाई होती है।

13.13 व्यावहारिक उदाहरण (Practical Examples of Information Retrieval)

सूचना पुनर्प्राप्ति की अवधारणाओं को बेहतर ढंग से समझने के लिए OPAC एवं ऑनलाइन डेटाबेस के कुछ **सरल एवं व्यावहारिक उदाहरण** प्रस्तुत किए जा रहे हैं।

(i) OPAC खोज (OPAC Search Example)

- **लेखक (Author):** *Ranganathan*
- **विषय (Subject):** *Library Science*

इस प्रकार की खोज से उपयोगकर्ता को संबंधित पुस्तकों एवं अन्य दस्तावेज़ों की सूची प्राप्त होती है।

(ii) डेटाबेस खोज (Database Search Example)

- **Boolean Query:**
  Library AND Automation

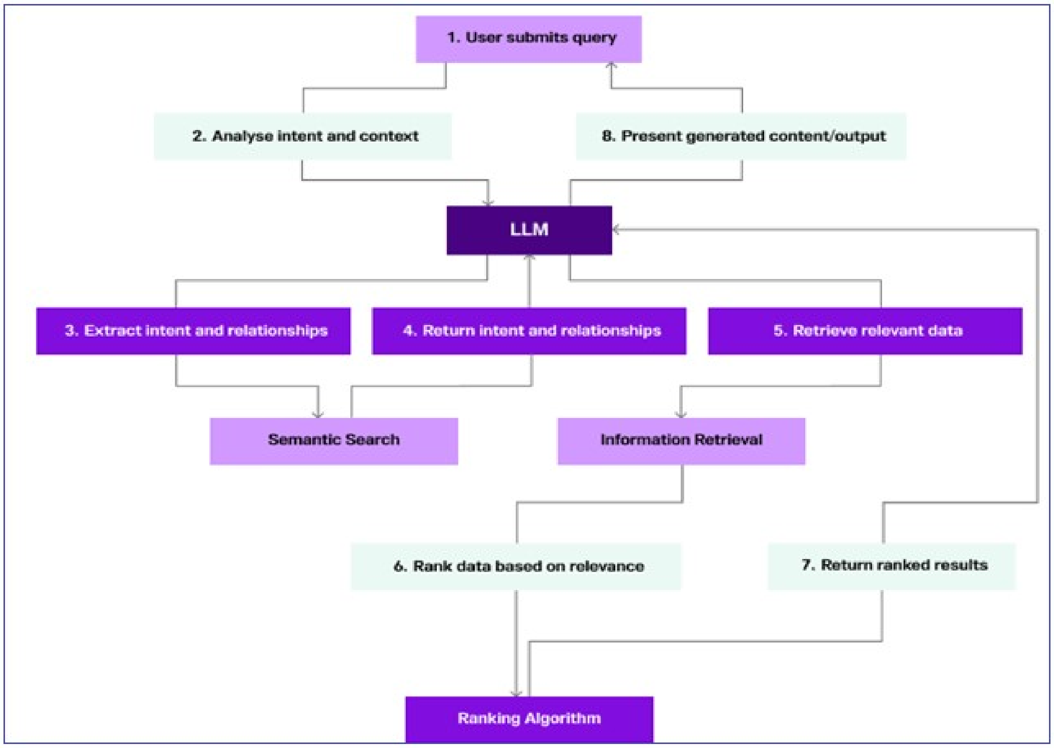
इस खोज में केवल वे दस्तावेज़ प्राप्त होंगे जिनमें *Library* और *Automation* दोनों शब्द मौजूद हों, जिससे परिणाम अधिक **सटीक (precise)** हो जाते हैं।

13.14 नवीन प्रवृत्तियाँ : AI आधारित IR सिस्टम (Emerging Trends: AI-based Information Retrieval Systems)

आधुनिक सूचना पुनर्प्राप्ति प्रणालियाँ अब **Artificial Intelligence (AI)** पर आधारित हो रही हैं, जिससे खोज अधिक **स्मार्ट (intelligent)**, **प्रासंगिक (relevant)** और **व्यक्तिगत (personalized)** बनती जा रही है।

AI आधारित IR की प्रमुख नवीन प्रवृत्तियाँ:

- **Semantic Search** – अर्थ आधारित खोज, न कि केवल शब्द आधारित
- **Natural Language Processing (NLP)** – प्राकृतिक भाषा में की गई queries को समझना
- **Recommendation Systems** – उपयोगकर्ता की रुचि के अनुसार सामग्री सुझाना



चित्र: AI आधारित सूचना पुनर्प्राप्ति प्रणाली की रूपरेखा

AI आधारित IR प्रणालियाँ उपयोगकर्ता अनुभव को बेहतर बनाती हैं और खोज परिणामों की गुणवत्ता में सुधार करती हैं।

13.15 सारांश (Summary)

इस इकाई में **सूचना पुनर्प्राप्ति (Information Retrieval)** की मूल अवधारणाओं से लेकर आधुनिक **AI आधारित सूचना पुनर्प्राप्ति प्रणालियों** तक का समग्र एवं व्यवस्थित अध्ययन प्रस्तुत किया गया है।
इकाई में सूचना पुनर्प्राप्ति प्रणाली के घटकों, प्रक्रिया, सूचना आवश्यकता एवं प्रासंगिकता, अनुक्रमण तकनीकों, खोज विधियों तथा पुनर्प्राप्ति की मापन विधियों (Precision, Recall, F-measure) को स्पष्ट रूप से समझाया गया है।
इसके अतिरिक्त डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति की भूमिका, उसकी सीमाएँ एवं चुनौतियाँ तथा **Semantic Search, NLP और Recommendation Systems** जैसी नवीन प्रवृत्तियों पर भी प्रकाश डाला गया है।
यह इकाई शिक्षार्थियों को सूचना पुनर्प्राप्ति की **सैद्धांतिक समझ (theoretical understanding)** के साथ-साथ **व्यावहारिक अनुप्रयोग (practical application)** के लिए भी सक्षम बनाती है।

13.16 शब्दावली (Glossary)

- **सूचना पुनर्प्राप्ति (Information Retrieval):** संग्रहीत सूचना-संसाधनों में से उपयोगकर्ता की आवश्यकता के अनुसार प्रासंगिक सूचना को खोजने की प्रक्रिया।
- **सूचना आवश्यकता (Information Need):** उपयोगकर्ता द्वारा अनुभव की गई वह बौद्धिक आवश्यकता, जिसके समाधान हेतु सूचना की आवश्यकता होती है।
- **प्रासंगिकता (Relevance):** खोज परिणाम और उपयोगकर्ता की सूचना आवश्यकता के बीच सामंजस्य का स्तर।
- **अनुक्रमण (Indexing):** दस्तावेज़ की विषयवस्तु को खोज-योग्य रूप में प्रस्तुत करने की प्रक्रिया।
- **खोज प्रश्न (Query):** उपयोगकर्ता द्वारा प्रणाली में प्रविष्ट किए गए शब्द या कथन, जिनके आधार पर खोज की जाती है।
- **की-वर्ड खोज (Keyword Search):** सीधे शब्दों के आधार पर की जाने वाली खोज विधि।

- **विषय आधारित खोज (Subject-based Search):** नियंत्रित शब्दावली या थिसॉरस के माध्यम से की जाने वाली खोज।
- **बूलियन खोज (Boolean Search):** AND, OR, NOT जैसे Boolean operators का उपयोग कर की जाने वाली खोज।
- **Precision:** प्राप्त दस्तावेज़ों में से प्रासंगिक दस्तावेज़ों का अनुपात।
- **Recall:** कुल उपलब्ध प्रासंगिक दस्तावेज़ों में से पुनर्प्राप्त किए गए दस्तावेज़ों का अनुपात।
- **F-measure (F-score):** Precision और Recall का संतुलित मापन।
- **डिजिटल पुस्तकालय (Digital Library):** डिजिटल रूप में संग्रहीत एवं नेटवर्क के माध्यम से उपलब्ध सूचना-संसाधनों का संग्रह।
- **Semantic Search:** अर्थ आधारित खोज तकनीक, जिसमें शब्दों के अर्थ को प्राथमिकता दी जाती है।
- **Natural Language Processing (NLP):** कंप्यूटर द्वारा मानव भाषा को समझने एवं संसाधित करने की तकनीक।
- **Recommendation System:** उपयोगकर्ता की रुचि एवं व्यवहार के आधार पर सामग्री सुझाने वाली प्रणाली।

13.17 निबंधात्मक प्रश्न (Essay Type Questions)

प्रश्न 1. सूचना पुनर्प्राप्ति प्रणाली (Information Retrieval System) के प्रमुख घटकों का वर्णन कीजिए।

उत्तर:
सूचना पुनर्प्राप्ति प्रणाली एक सुव्यवस्थित तंत्र है, जिसका उद्देश्य उपयोगकर्ता की सूचना आवश्यकता के अनुरूप प्रासंगिक सूचना उपलब्ध कराना है। इसके प्रमुख घटक हैं—
(1) सूचना संसाधन, जिनमें पुस्तकें, जर्नल, शोध-प्रबंध, ई-संसाधन एवं मल्टीमीडिया सामग्री सम्मिलित होती है।
(2) उपयोगकर्ता, जो अपनी सूचना आवश्यकता के अनुसार खोज करता है।
(3) खोज इंटरफेस, जिसके माध्यम से उपयोगकर्ता और प्रणाली के बीच संवाद होता है।
(4) पुनर्प्राप्ति तंत्र, जिसमें अनुक्रमण, खोज एल्गोरिद्म तथा रैंकिंग प्रणाली शामिल होती है।
इन घटकों का समन्वित कार्य ही सूचना पुनर्प्राप्ति प्रणाली की प्रभावशीलता को सुनिश्चित करता है।

प्रश्न 2. सूचना पुनर्प्राप्ति प्रक्रिया को चरणबद्ध रूप में समझाइए।

उत्तर:
सूचना पुनर्प्राप्ति प्रक्रिया एक क्रमिक एवं तार्किक प्रक्रिया है। इसमें सर्वप्रथम उपयोगकर्ता अपनी सूचना आवश्यकता की पहचान करता है। इसके पश्चात खोज प्रश्न (query) का निर्माण किया जाता है। प्रणाली अनुक्रमण एवं मिलान के माध्यम से उपयुक्त दस्तावेज़ों की पहचान कर प्रासंगिक परिणाम प्रस्तुत करती है। अंत में उपयोगकर्ता प्राप्त परिणामों का मूल्यांकन करता है। यह प्रक्रिया आवश्यकतानुसार दोहराई भी जा सकती है।

प्रश्न 3. Precision एवं Recall को उदाहरण सहित स्पष्ट कीजिए।

उत्तर:
Precision वह अनुपात है, जो यह दर्शाता है कि प्राप्त दस्तावेज़ों में से कितने वास्तव में प्रासंगिक हैं।
Recall यह दर्शाता है कि कुल उपलब्ध प्रासंगिक दस्तावेज़ों में से कितने पुनर्प्राप्त किए गए हैं।
उदाहरणतः यदि 10 दस्तावेज़ प्राप्त हुए, जिनमें 6 प्रासंगिक हैं, तो Precision = 6/10 होगा।
यदि कुल प्रासंगिक दस्तावेज़ 12 थे और 6 प्राप्त हुए, तो Recall = 6/12 होगा।

प्रश्न 4. डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति की भूमिका स्पष्ट कीजिए।

उत्तर:
डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति की भूमिका अत्यंत महत्वपूर्ण है, क्योंकि डिजिटल संग्रह विशाल, विविध एवं बहुरूपात्मक होता है। प्रभावी IR प्रणाली उपयोगकर्ताओं को सटीक, शीघ्र एवं प्रासंगिक सूचना उपलब्ध कराकर ज्ञान-प्राप्ति की प्रक्रिया को सरल बनाती है। यह उपयोगकर्ता अनुभव में सुधार करती है तथा डिजिटल संसाधनों के प्रभावी उपयोग को सुनिश्चित करती है।

प्रश्न 5. AI आधारित सूचना पुनर्प्राप्ति प्रणालियों की प्रमुख विशेषताओं का वर्णन कीजिए।

उत्तर:
AI आधारित सूचना पुनर्प्राप्ति प्रणालियाँ Semantic Search, Natural Language Processing तथा Recommendation Systems जैसी तकनीकों पर आधारित होती हैं। ये प्रणालियाँ उपयोगकर्ता की खोज मंशा को समझकर अधिक सटीक, प्रासंगिक एवं व्यक्तिगत खोज परिणाम प्रदान करती हैं, जिससे खोज की गुणवत्ता में उल्लेखनीय सुधार होता है।

13.18 लघु प्रश्न (Short Answer Questions)

1. सूचना पुनर्प्राप्ति क्या है?
2. सूचना आवश्यकता से आप क्या समझते हैं?
3. Boolean Search क्या है?

4. अनुक्रमण का क्या महत्व है?
5. F-measure का प्रयोग क्यों किया जाता है?

13.19 बहुविकल्पीय प्रश्न (MCQs) – उत्तर सहित

i. सूचना पुनर्प्राप्ति का मुख्य उद्देश्य है—
a) सूचना का संग्रह
b) सूचना का संरक्षण
c) प्रासंगिक सूचना की खोज
d) सूचना का प्रकाशन

ii. Precision किसे मापता है?
a) कुल दस्तावेज़
b) **प्राप्त दस्तावेज़ों में प्रासंगिक दस्तावेज़ों का अनुपात**
c) संग्रह का आकार
d) खोज समय

iii. Recall का संबंध किससे है?
a) प्राप्त दस्तावेज़
b) **कुल प्रासंगिक दस्तावेज़**
c) उपयोगकर्ता
d) इंटरफेस

iv. Boolean operator AND का प्रभाव होता है—
a) खोज को विस्तृत करना
b) **खोज को संकुचित करना**
c) खोज रोकना
d) खोज हटाना

v. OPAC का पूर्ण रूप है—
a) **Online Public Access Catalogue**
b) Open Public Access Control
c) Online Program Access Catalogue
d) Open Program Access Control

vi. अनुक्रमण का मुख्य उद्देश्य है—
a) संग्रह बढ़ाना
b) **खोज को सरल बनाना**

c) संग्रह हटाना
d) डेटा एन्क्रिप्शन

vii. F-measure किसका संतुलन दर्शाता है?
a) Precision और Speed
b) Recall और Speed
c) Precision और Recall
d) Indexing और Searching

viii. Semantic Search आधारित होती है—
a) शब्दों की संख्या पर
b) अर्थ (Meaning) पर
c) वर्णमाला पर
d) फ़ाइल आकार पर

ix. NLP का पूर्ण रूप है—
a) Natural Language Program
b) Network Language Processing
c) Natural Language Processing ✔
d) New Language Protocol

x. Recommendation System का उद्देश्य है—
a) खोज रोकना
b) उपयोगकर्ता को उपयुक्त सामग्री सुझाना ✔
c) संग्रह हटाना
d) डेटा संग्रह

13.20 संदर्भ ग्रंथ सूची (References)

1. शर्मा, आर. एन. – सूचना पुनर्प्राप्ति प्रणाली
2. मिश्रा, एस. पी. – डिजिटल पुस्तकालय
3. सिंह, के. एल. – सूचना विज्ञान के सिद्धांत
4. वर्मा, बी. डी. – आधुनिक पुस्तकालय सेवाएँ
5. कुमार, अजय – सूचना संगठन एवं पुनर्प्राप्ति
6. जोशी, एम. के. – पुस्तकालय एवं सूचना विज्ञान
7. त्रिपाठी, आर. – डिजिटल युग में पुस्तकालय
8. पांडेय, एस. – सूचना प्रबंधन
9. अग्रवाल, पी. – ज्ञान संगठन

10. चौधरी, ए. – सूचना पुनर्प्राप्ति तकनीकें
11. Chowdhury, G. G. (2010). Introduction to Modern Information Retrieval. Facet Publishing.
12. Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern Information Retrieval. Addison-Wesley.
13. Lancaster, F. W. (2003). Information Retrieval Systems. Academic Press.
14. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
15. Salton, G., & McGill, M. J. (1983). Introduction to Modern Information Retrieval. McGraw-Hill.
16. Croft, W. B., Metzler, D., & Strohman, T. (2010). Search Engines: Information Retrieval in Practice. Pearson.
17. Rowley, J., & Farrow, J. (2000). Organizing Knowledge. Ashgate.
18. Chowdhury, G. G., & Chowdhury, S. (2007). Organizing Information. Facet Publishing.
19. Hearst, M. (2009). Search User Interfaces. Cambridge University Press.
20. Lesk, M. (2004). Understanding Digital Libraries. Morgan Kaufmann.

इकाई–14 : सूचना पुनर्प्राप्ति मॉडल एवं उनके अनुप्रयोग (Information Retrieval – Models and Their Applications)

इकाई की रूपरेखा

14.1 प्रस्तावना
14.2 उद्देश्य
14.3 सूचना पुनर्प्राप्ति मॉडल : अवधारणा
14.4 सूचना पुनर्प्राप्ति मॉडल के प्रकार
14.5 बूलियन मॉडल **(Boolean Model)**
14.5.1 विशेषताएँ
14.5.2 लाभ एवं सीमाएँ
14.6 वेक्टर स्पेस मॉडल **(Vector Space Model)**
14.6.1 अवधारणा
14.6.2 समानता मापन (Cosine Similarity)
14.7 प्रायिकता मॉडल **(Probabilistic Model)**
14.7.1 मूल सिद्धांत
14.7.2 उपयोग
14.8 भाषा मॉडल **(Language Model)**
14.9 संरचित एवं असंरचित डेटा में **IR** मॉडल
14.10 विभिन्न मॉडल्स की तुलनात्मक विवेचना
14.11 आधुनिक सूचना पुनर्प्राप्ति मॉडल
14.11.1 फजी मॉडल
14.11.2 न्यूरल नेटवर्क आधारित मॉडल
14.12 सूचना पुनर्प्राप्ति मॉडल के अनुप्रयोग क्षेत्र
14.12.1 डिजिटल पुस्तकालय
14.12.2 वेब खोज इंजन
14.12.3 अकादमिक डेटाबेस
14.13 व्यावहारिक उदाहरण: मॉडल आधारित खोज प्रणाली
14.14 भविष्य की दिशा एवं शोध संभावनाएँ
14.15 सारांश
14.16 शब्दावली

14.17 निबंधात्मक प्रश्न
14.18 लघु प्रश्न एवं MCQ
14.19 अभ्यास प्रश्नों के उत्तर
14.20 संदर्भ ग्रंथ सूची

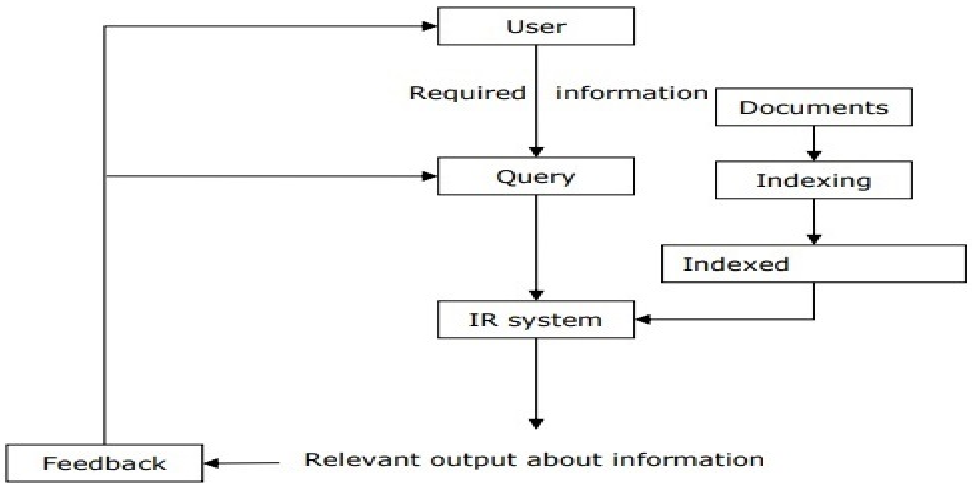
14.1 प्रस्तावना (Introduction)

सूचना पुनर्प्राप्ति (Information Retrieval – IR) केवल सूचना खोजने की एक साधारण प्रक्रिया नहीं है, बल्कि यह एक **वैज्ञानिक, तार्किक एवं मॉडल-आधारित प्रणाली (model-based system)** है।
जैसे-जैसे सूचना संसाधनों (Information Resources) की मात्रा और विविधता में वृद्धि हुई है, वैसे-वैसे **प्रभावी एवं सटीक सूचना खोज (effective and precise information search)** के लिए विभिन्न **सूचना पुनर्प्राप्ति मॉडल (IR Models)** विकसित किए गए हैं।

सूचना पुनर्प्राप्ति मॉडल यह स्पष्ट करते हैं कि—

- सूचना को किस प्रकार **represent (document representation)** किया जाए,
- उपयोगकर्ता की **query को कैसे process** किया जाए, तथा
- दस्तावेज़ों की **प्रासंगिकता (relevance determination)** किस आधार पर निर्धारित की जाए।

अतः IR models सूचना पुनर्प्राप्ति प्रणाली की **कार्य-प्रणाली (working mechanism)** को समझने का मूल आधार प्रदान करते हैं।



चित्र 14.1 : सूचना पुनर्प्राप्ति मॉडल का सामान्य संरचनात्मक ढाँच

14.2 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात शिक्षार्थी—

- सूचना पुनर्प्राप्ति मॉडल (Information Retrieval Models) की **अवधारणा** को स्पष्ट रूप से समझ सकेंगे
- विभिन्न IR models का **वर्गीकरण (classification)** कर सकेंगे
- Boolean, Vector Space एवं Probabilistic models का **तुलनात्मक अध्ययन (comparative analysis)** कर सकेंगे
- आधुनिक **AI-आधारित सूचना पुनर्प्राप्ति मॉडलों** की भूमिका को समझ सकेंगे
- डिजिटल पुस्तकालयों, वेब खोज इंजनों एवं अकादमिक डेटाबेस में IR models के **व्यावहारिक अनुप्रयोगों (applications)** की पहचान कर सकेंगे

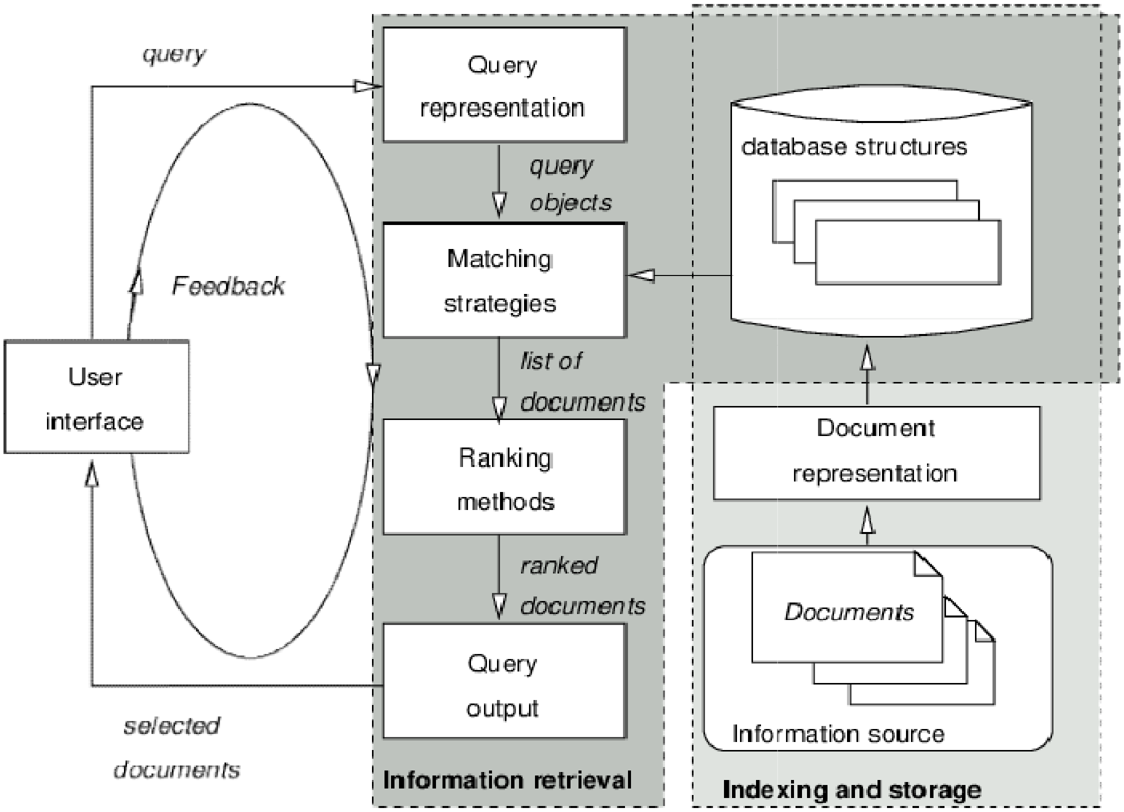
14.3 सूचना पुनर्प्राप्ति मॉडल : अवधारणा(Concept of Information Retrieval Models)

सूचना पुनर्प्राप्ति मॉडल (Information Retrieval Models) वह **सैद्धांतिक ढाँचा (theoretical framework)** है, जिसके माध्यम से यह समझा जाता है कि **सूचना संसाधन (information resources)**, **उपयोगकर्ता की query**, तथा **retrieval mechanism** आपस में किस प्रकार अंतःक्रिया (interaction) करते हैं।

सरल शब्दों में—

IR Model यह निर्धारित करता है कि "क्या खोजा जाए (what to search), कैसे खोजा जाए (how to search) और किस आधार पर परिणामों को क्रमबद्ध किया जाए (how to rank results)।"

ये मॉडल सूचना को **represent**, **match** तथा **evaluate** करने की एक व्यवस्थित पद्धति प्रदान करते हैं और किसी भी सूचना पुनर्प्राप्ति प्रणाली की कार्य-प्रणाली को समझने का आधार बनते हैं।



चित्र 14.1 : सूचना पुनर्प्राप्ति मॉडल का सामान्य संरचनात्मक ढाँचा
Figure 14.1: General Architecture of an Information Retrieval **Model**

14.4 सूचना पुनर्प्राप्ति मॉडल के प्रकार (Types of Information Retrieval Models)

सूचना पुनर्प्राप्ति मॉडल को उनके **matching approach**, **relevance determination** तथा **result presentation** के आधार पर निम्नलिखित प्रमुख वर्गों में विभाजित किया जा सकता है—

1. **Boolean Model** – तार्किक (logic-based) मॉडल
2. **Vector Space Model** – वेक्टर एवं समानता आधारित मॉडल
3. **Probabilistic Model** – प्रायिकता (probability) आधारित मॉडल
4. **Language Model** – भाषा-आधारित (language-based) मॉडल
5. **आधुनिक मॉडल (Modern Models)** –
   जैसे **Fuzzy Models** एवं **Neural Network / AI-based Models**

ये सभी मॉडल विभिन्न प्रकार की सूचना आवश्यकताओं और अनुप्रयोगों के लिए उपयोगी हैं, जैसे—डिजिटल पुस्तकालय, वेब खोज इंजन तथा अकादमिक डेटाबेस।

14.5 बूलियन मॉडल (Boolean Model)

बूलियन मॉडल (Boolean Model) सूचना पुनर्प्राप्ति का **सबसे प्रारंभिक एवं पारंपरिक मॉडल (classical IR model)** है। यह मॉडल **Boolean logic** पर आधारित होता है और दस्तावेज़ों तथा उपयोगकर्ता की query के बीच **exact matching** को प्राथमिकता देता है।

इस मॉडल में दस्तावेज़ या तो **प्रासंगिक (relevant)** होते हैं या **अप्रासंगिक (non-relevant)** — बीच की कोई स्थिति नहीं होती।

14.5.1 विशेषताएँ (Features)

- Boolean logic **(AND, OR, NOT)** पर आधारित
- Query और document के बीच **exact match**
- **Relevance ranking** की सुविधा उपलब्ध नहीं
- परिणाम **binary form** में प्राप्त होते हैं (Yes / No)

14.5.2 लाभ एवं सीमाएँ (Advantages and Limitations)

लाभ (Advantages):

- सरल एवं स्पष्ट (simple and easy to understand)
- **OPAC** एवं प्रारंभिक **bibliographic databases** में उपयोगी
- Controlled vocabulary आधारित खोज में प्रभावी

सीमाएँ (Limitations):

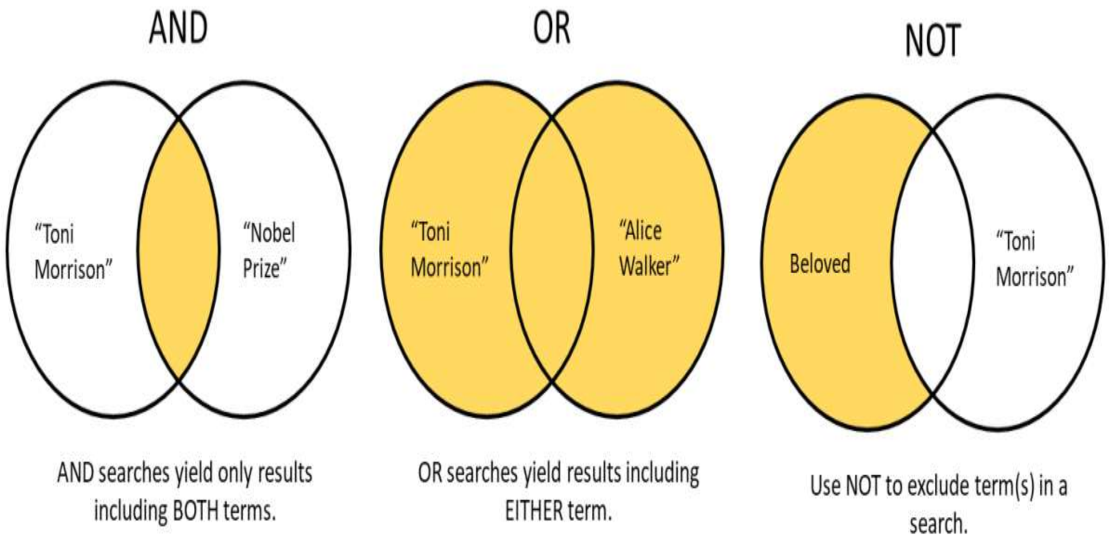
- **Partial relevance** या relevance degree को प्रदर्शित नहीं करता
- Result ranking (क्रमबद्ध परिणाम) उपलब्ध नहीं
- Large databases में बहुत अधिक या बहुत कम परिणाम दे सकता है

उदाहरण (Example)

Boolean Query:
Library AND Automation

इस query में केवल वही दस्तावेज़ प्राप्त होंगे जिनमें **Library** और **Automation** दोनों शब्द उपस्थित हों।

Suggested Image (Highly Recommended)



चित्र 14.2 : बूलियन खोज मॉडल (AND, OR, NOT) का वेन आरेख
(Figure 14.2: Venn Diagram of Boolean Search Model)

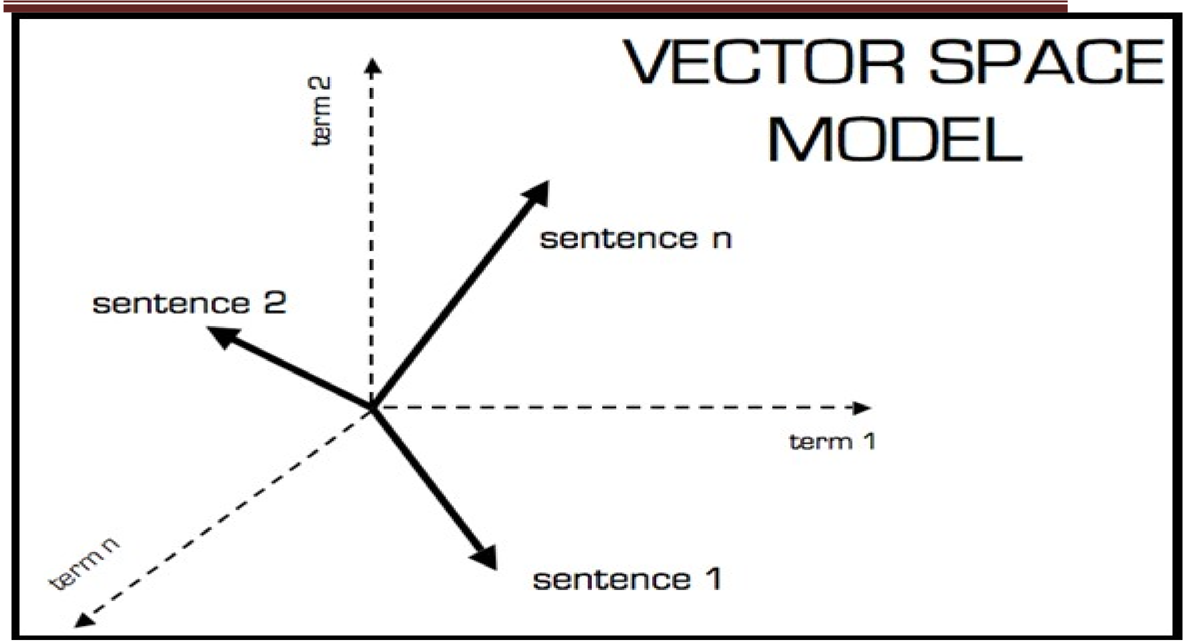
इस प्रकार **Boolean model** में परिणामों का चयन पूर्णतः तार्किक शर्तों **(logical conditions)** पर आधारित होता है।

14.6 वेक्टर स्पेस मॉडल (Vector Space Model – VSM):

14.6.1 अवधारणा (Concept): वेक्टर स्पेस मॉडल (Vector Space Model) एक **similarity-based Information Retrieval model** है, जिसमें—

- **Document** और **User Query** को
- **Multi-dimensional vectors** (बहु-आयामी वेक्टर) के रूप में
- Mathematical space में प्रस्तुत किया जाता है।

प्रत्येक **dimension** एक **term/keyword** को दर्शाती है।



सरल शब्दों में :
Boolean Model "match है या नहीं" देखता है,
जबकि **VSM यह देखता है कि *कितना match है*।**

Term Weighting (Very important)

VSM में हर शब्द समान महत्वपूर्ण नहीं होता।
इसलिए **Term Weighting** प्रयोग की जाती है, जैसे:

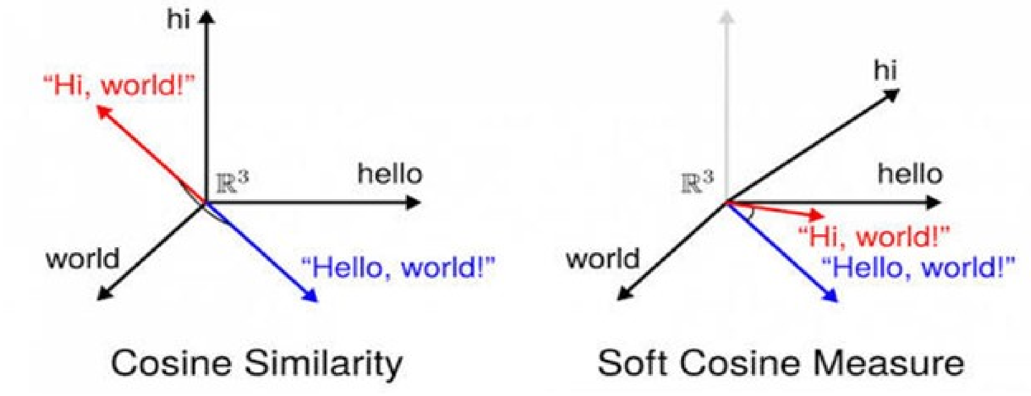
- **TF (Term Frequency)** → शब्द document में कितनी बार आया
- **IDF (Inverse Document Frequency)** → शब्द पूरे database में कितना rare है

Rare लेकिन meaningful शब्दों को **ज्यादा weight** मिलता है।

14.6.2 समानता मापन (Cosine Similarity)

VSM में similarity प्रायः **Cosine Similarity** से मापी जाती है।

- Query vector और Document vector के बीच **angle (θ)** निकाला जाता है
- Angle जितना **छोटा** → similarity उतनी **अधिक**
- Value सामान्यतः **0 से 1** के बीच होती है



उदाहरण (Library-centric)

Query: "Digital Library Software"

- **Document A:** "KOHA is an open-source library automation software"
- **Document B:** "History of ancient libraries"

✔Document A में terms overlap ज़्यादा → **higher similarity** → **top rank**

VSM –लाभ

- Partial matching संभव
- Result ranking उपलब्ध
- Modern search systems में widely used

14.7 प्रायिकता मॉडल (Probabilistic Model)

14.7.1 मूल सिद्धांत (Basic Principle)

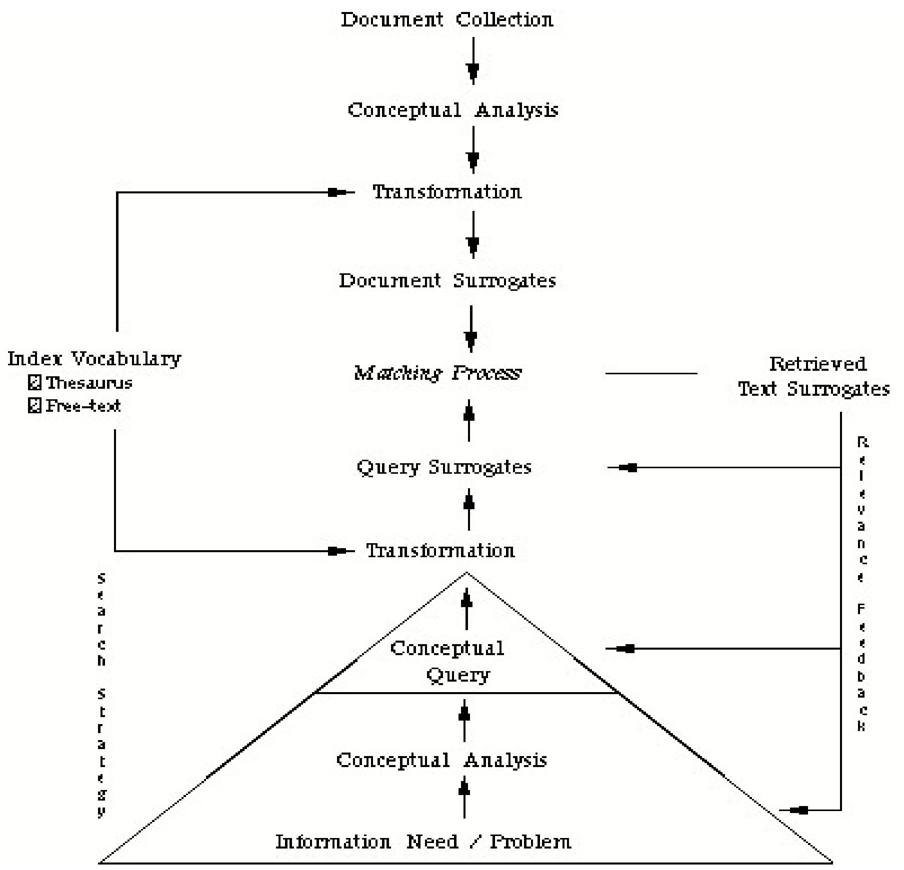
प्रायिकता मॉडल इस विचार पर आधारित है कि—

हर document के relevant होने की एक probability होती है।

System यह अनुमान लगाता है:

- कौन-सा document user की query को
- **सबसे अधिक satisfy** करेगा

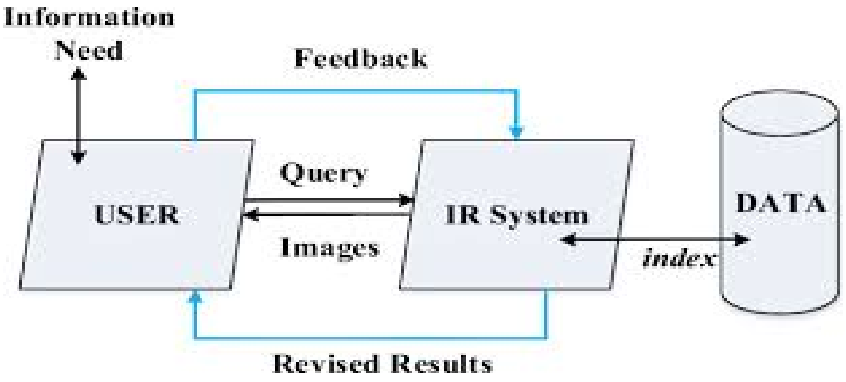
✓ जिस document की probability ज़्यादा → **rank ऊपर**



Probability Ranking Principle (PRP) –

PRP के अनुसार:

Documents को **probability of relevance** के **decreasing order** में rank करना सबसे प्रभावी retrieval देता है।



14.7.2 उपयोग (Applications)

- Modern Search Engines
- Relevance Feedback Systems
- Large Academic Databases

Example

Query: "Library Automation Software"

- Documents जिनमें *KOHA, DSpace, Open Source* जैसे terms हैं
- और जिन्हें users पहले relevant मान चुके हैं

✔ उनकी relevance probability ज़्यादा → **top results**

14.8 भाषा मॉडल (Language Model) (Language Model in Information Retrieval):

भाषा मॉडल (Language Model) सूचना पुनर्प्राप्ति का एक **आधुनिक एवं प्रायिकता-आधारित मॉडल (probabilistic-based IR model)** है। इस मॉडल में यह माना जाता है कि प्रत्येक दस्तावेज़ (document) एक **भाषा का नमूना (language sample)** है और उपयोगकर्ता की query भी उसी भाषा का एक छोटा रूप होती है।

इस मॉडल का मुख्य उद्देश्य यह निर्धारित करना होता है कि—

कौन-सा document उपयोगकर्ता की query को उत्पन्न (generate) करने की सबसे अधिक संभावना रखता है।

अर्थात्, जिस document से query उत्पन्न होने की संभावना अधिक होती है, उसे उतना ही अधिक **प्रासंगिक (relevant)** माना जाता है।

मुख्य विशेषताएँ (Key Features)

- **Probability-based approach** पर आधारित
- Query likelihood का आकलन करता है
- **Relevance ranking** की सुविधा उपलब्ध
- Modern IR systems में व्यापक उपयोग

उपयोग (Applications)

- **Modern search engines**
- **Statistical language processing systems**
- **AI-based and semantic search environments**

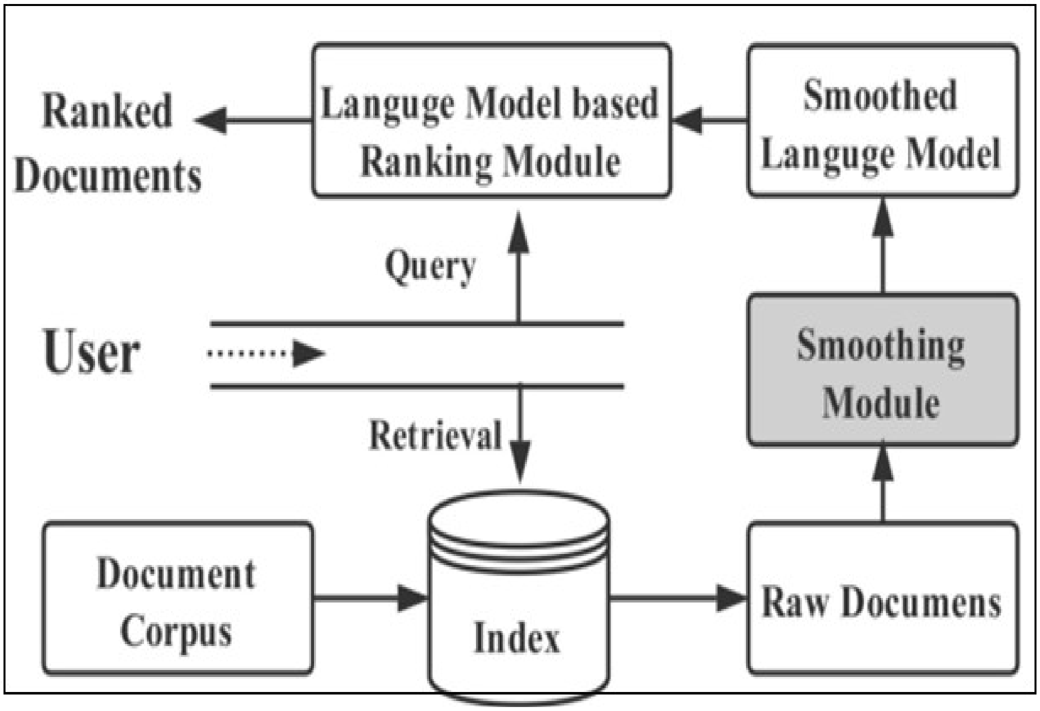


Fig.: Language Model–Based Information Retrieval System

Explaination: यह चित्र दिखाता है कि **यूज़र की query** को भाषा मॉडल और **smoothing तकनीक** की मदद से प्रोसेस करके दस्तावेज़ों को **महत्व के अनुसार क्रम (rank)** में दिखाया जाता है।

14.9 संरचित एवं असंरचित डेटा में IR मॉडल

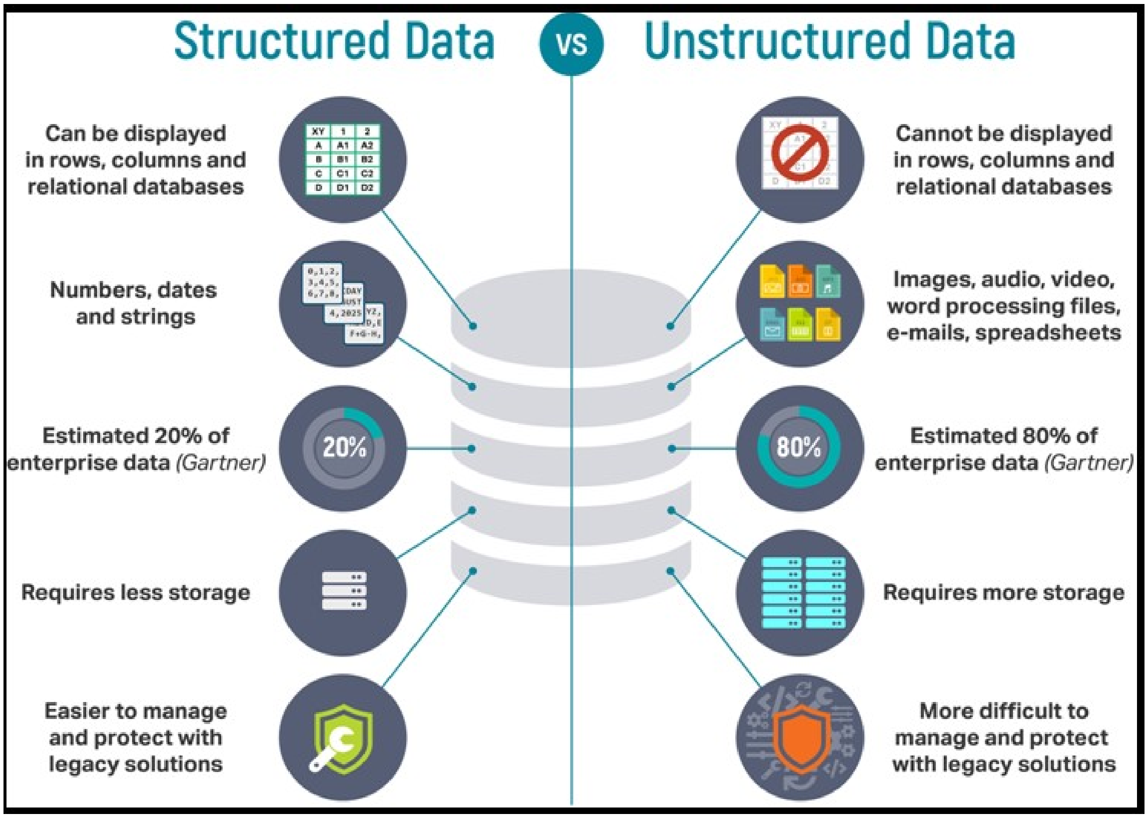
सूचना पुनर्प्राप्ति मॉडल का चयन डेटा की प्रकृति (nature of data) पर निर्भर करता है।

- **Structured Data**
  जैसे: Databases, Metadata records, OPAC entries
  → Boolean Model, Exact matching models अधिक उपयुक्त होते हैं।

- **Unstructured Data**
  जैसे: Text documents, Web pages, Multimedia content

→ Vector Space Model, Probabilistic Model, Language Model अधिक प्रभावी होते हैं।

इस प्रकार, different IR models का उपयोग data structure के अनुसार किया जाता है।



Structured vs Unstructured Data flow diagram

14.10 विभिन्न मॉडल्स की तुलनात्मक विवेचना

Model	Matching Type	Ranking	Complexity
Boolean Model	Exact	No	Low
Vector Space Model	Partial	Yes	Medium
Probabilistic Model	Probability-based	Yes	High
Language Model	Statistical	Yes	High

Comparison table

14.11 आधुनिक सूचना पुनर्प्राप्ति मॉडल(Modern Information Retrieval Models)

परंपरागत IR models की सीमाओं को दूर करने के लिए आधुनिक सूचना पुनर्प्राप्ति मॉडल विकसित किए गए हैं, जो **vagueness, uncertainty और user intent** को बेहतर ढंग से संभालते हैं

14.11.1 फजी मॉडल (Fuzzy Model)

फजी मॉडल (Fuzzy IR Model) **fuzzy logic** पर आधारित होता है, जिसमें किसी दस्तावेज़ की प्रासंगिकता (relevance) को **degree या scale (0–1)** के रूप में मापा जाता है।

मुख्य विशेषताएँ (Key Features):

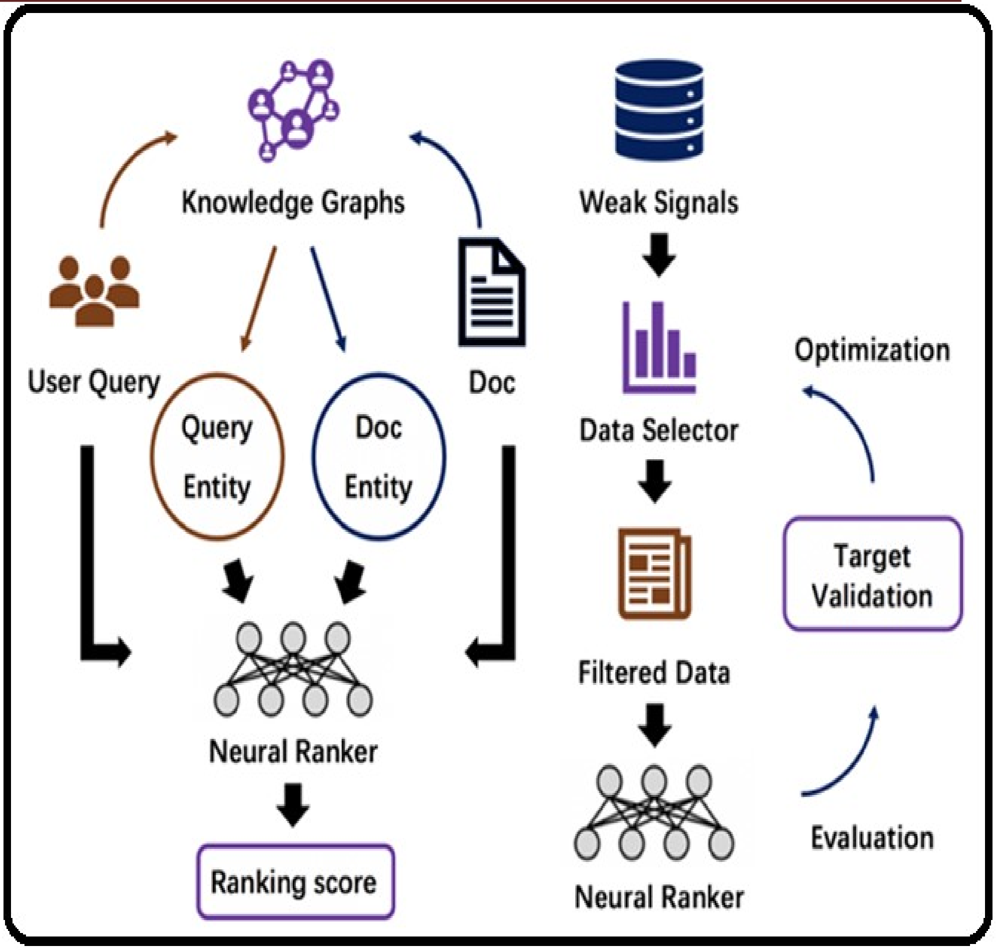
- Relevance को **binary (Yes/No)** के बजाय **degree of relevance** में दर्शाता है
- Partial matching को support करता है
- Human thinking pattern के अधिक निकट

A Fuzzy IR Model allows documents to be relevant to varying degrees rather than treating relevance as absolute.

उदाहरण:
यदि query है **"Digital Library Automation"**, तो कोई document **0.8 relevance** और दूसरा **0.5 relevance** score प्राप्त कर सकता है।

14.11.2 न्यूरल नेटवर्क आधारित मॉडल (Neural Network Based IR Models)

न्यूरल नेटवर्क आधारित मॉडल **Machine Learning (ML)** एवं **Deep Learning (DL)** तकनीकों पर आधारित होते हैं। ये मॉडल user query और document content के बीच **semantic relationship** को समझते हैं।



Neural Network आधारित मॉडल

मुख्य विशेषताएँ (Key Features):

- Learning-based (data से सीखने की क्षमता)
- Semantic understanding और context awareness
- Highly accurate ranking results

Applications:

- Google Search
- Semantic Scholar
- Recommendation systems

Neural Network based IR models learn complex patterns from data to improve relevance ranking.

14.12 सूचना पुनर्प्राप्ति मॉडल के अनुप्रयोग क्षेत्र (Applications of IR Models):

सूचना पुनर्प्राप्ति मॉडल विभिन्न digital information environments में व्यापक रूप से प्रयुक्त होते हैं।

14.12.1 डिजिटल पुस्तकालय (Digital Libraries)

- Large digital collections की efficient retrieval
- Metadata + full-text search

Examples:
DSpace, Greenstone, EPrints

14.12.2 वेब खोज इंजन (Web Search Engines)

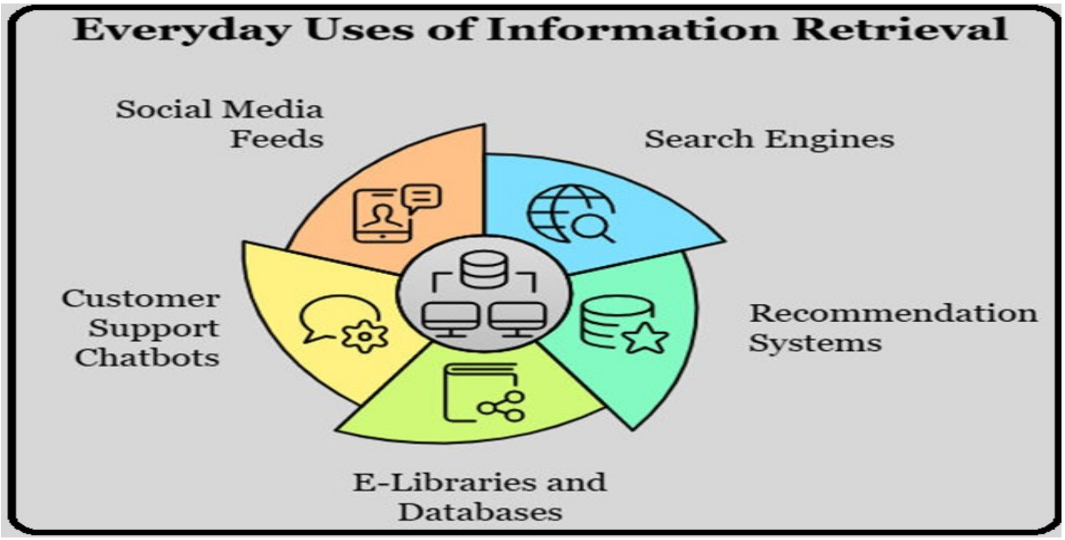
- Billions of web pages में से relevant information retrieval
- Ranking algorithms का प्रयोग

Examples:

Google, Bing

14.12.3 अकादमिक डेटाबेस (Academic Databases)

- Research-oriented precise searching
- Citation-based ranking



Examples:
Scopus, Web of Science

4.13 व्यावहारिक उदाहरण (Practical Example)

उदाहरण:
किसी **Academic Database** (जैसे *Scopus* या *Web of Science*) में जब उपयोगकर्ता Digital Library Systems विषय पर खोज करता है, तब **Vector Space Model (VSM)** के माध्यम से documents और query को **multi-dimensional vectors** के रूप में represent किया जाता है।

इसके पश्चात **Cosine Similarity** के आधार पर documents को **relevance score** प्रदान किया जाता है और उपयोगकर्ता को **ranked search results** (most relevant first) प्राप्त होते हैं।

- ✓ यह उदाहरण दर्शाता है कि **Vector Space Model** कैसे **partial matching** और **ranking** को संभव बनाता है।

14.14 भविष्य की दिशा एवं शोध संभावनाएँ (Future Directions and Research Opportunities)

सूचना पुनर्प्राप्ति (Information Retrieval) का भविष्य **Artificial Intelligence (AI)** और **Machine Learning (ML)** के साथ तेज़ी से विकसित हो रहा है। प्रमुख शोध दिशाएँ निम्नलिखित हैं—

• **AI-based Semantic Retrieval** – जहाँ system शब्दों के बजाय **meaning (semantics)** को समझकर खोज करता है।

• **Context-aware Search** – जिसमें user का context, location, past behaviour और intent
search results को प्रभावित करता है।

• **Personalized Information Retrieval Systems** – जहाँ प्रत्येक उपयोगकर्ता को उसकी preferences और interests के अनुसार customized results मिलते हैं।

✓ ये प्रवृत्तियाँ **next-generation search systems** और **smart digital libraries** के विकास में महत्वपूर्ण भूमिका निभाएँगी।

14.15 सारांश (Summary)

इस इकाई में **सूचना पुनर्प्राप्ति मॉडलों (Information Retrieval Models)** की अवधारणा, प्रकार, कार्य-प्रणाली तथा उनके व्यावहारिक अनुप्रयोगों का समग्र अध्ययन प्रस्तुत किया गया है।
Boolean Model जैसे पारंपरिक मॉडलों से प्रारंभ करते हुए **Vector Space Model** एवं **Probabilistic Model** की कार्यप्रणाली को उदाहरण सहित स्पष्ट किया गया है। साथ ही, आधुनिक **AI-आधारित एवं Neural Network आधारित IR Models** ने सूचना खोज (information searching) को अधिक **effective, accurate, ranked** एवं **user-centric** बना दिया है।

डिजिटल पुस्तकालयों, वेब खोज इंजनों तथा अकादमिक डेटाबेस में इन मॉडलों की भूमिका आधुनिक सूचना सेवाओं की आधारशिला के रूप में उभरकर सामने आती है।

14.16 शब्दावली (Glossary)

English Term		हिन्दी अर्थ / व्याख्या
Information Retrieval(IR):		संग्रहीत सूचना में से प्रासंगिक सूचना को खोजने की प्रक्रिया
Relevance	:	उपयोगकर्ता की सूचना आवश्यकता से परिणाम का सामंजस्य
Query	:	उपयोगकर्ता द्वारा दिया गया खोज प्रश्न
Document Corpus	:	खोज हेतु उपलब्ध दस्तावेज़ों का संग्रह
Boolean Model	:	तार्किक ऑपरेटर (AND, OR, NOT) पर आधारित IR मॉडल
Vector Space Model (VSM):		दस्तावेज़ एवं query को वेक्टर के रूप में निरूपित करने वाला मॉडल
Cosine Similarity	:	दो वेक्टरों के बीच समानता मापने की गणितीय विधि
Probabilistic Model	:	प्रासंगिकता की संभावना (probability) पर आधारित IR मॉडल
Ranking	:	खोज परिणामों को प्रासंगिकता के अनुसार क्रमबद्ध करना
Indexing	:	दस्तावेज़ों को खोज-योग्य रूप में व्यवस्थित करने की प्रक्रिया
Relevance Feedback	:	उपयोगकर्ता प्रतिक्रिया के आधार पर खोज सुधार तकनीक
Neural Network Model	:	मशीन लर्निंग/डीप लर्निंग पर आधारित आधुनिक IR मॉडल
Semantic Search	:	अर्थ (meaning) आधारित खोज प्रणाली
User-centric Search	:	उपयोगकर्ता व्यवहार एवं आवश्यकता पर आधारित खोज

14.17 निबंधात्मक प्रश्न – उत्तर सहित (Long Answer Type)

प्रश्न 1. सूचना पुनर्प्राप्ति मॉडल (Information Retrieval Models) की अवधारणा स्पष्ट कीजिए।

उत्तर: सूचना पुनर्प्राप्ति मॉडल वह **theoretical framework** है जो यह स्पष्ट करता है कि सूचना संसाधनों (documents), उपयोगकर्ता की खोज-प्रश्न (query) तथा पुनर्प्राप्ति तंत्र (retrieval mechanism) के बीच किस प्रकार अंतःक्रिया होती है। IR मॉडल यह निर्धारित करता है कि दस्तावेज़ों को किस रूप में प्रस्तुत किया जाए

(representation), query का विश्लेषण कैसे किया जाए तथा प्रासंगिकता (relevance) के आधार पर परिणामों को कैसे चुना एवं क्रमबद्ध किया जाए।

Boolean Model, Vector Space Model तथा Probabilistic Model प्रमुख IR मॉडल हैं, जो सूचना खोज की प्रभावशीलता बढ़ाने में सहायक होते हैं। आधुनिक समय में AI एवं Neural Network आधारित मॉडल भी व्यापक रूप से उपयोग किए जा रहे हैं।

प्रश्न 2. Boolean Model एवं Vector Space Model की तुलना कीजिए।

उत्तर:

आधार	**Boolean Model**	**Vector Space Model**
Matching	Exact matching	Partial matching
Representation	Binary (Yes/No)	Vector based
Ranking	उपलब्ध नहीं	उपलब्ध
Flexibility	कम	अधिक
उपयोग	OPAC, प्रारंभिक databases	Modern search systems

Boolean Model सरल है परंतु relevance ranking नहीं देता, जबकि Vector Space Model cosine similarity के माध्यम से documents को ranked रूप में प्रस्तुत करता है।

प्रश्न 3. डिजिटल पुस्तकालयों में सूचना पुनर्प्राप्ति मॉडलों की भूमिका स्पष्ट कीजिए।

उत्तर:
डिजिटल पुस्तकालयों में संसाधनों की संख्या अत्यधिक होती है, इसलिए प्रभावी सूचना खोज के लिए IR models की भूमिका अत्यंत महत्वपूर्ण है।
Vector Space Model और Probabilistic Model डिजिटल पुस्तकालयों में उपयोगकर्ताओं को ranked search results प्रदान करते हैं।
DSpace, Greenstone और EPrints जैसे digital library platforms आधुनिक IR models का प्रयोग कर उपयोगकर्ता-केन्द्रित खोज सुविधा प्रदान करते हैं।

प्रश्न 4. Probabilistic Model की कार्य-प्रणाली समझाइए।

उत्तर:
Probabilistic Model इस सिद्धांत पर आधारित है कि प्रत्येक दस्तावेज़ के प्रासंगिक होने की एक **probability** होती है।
यह मॉडल उपयोगकर्ता की query और पूर्व अनुभव (relevance feedback) के आधार पर यह

अनुमान लगाता है कि कौन-सा दस्तावेज़ अधिक प्रासंगिक है। इस मॉडल में relevance को binary नहीं बल्कि probability के रूप में देखा जाता है, जिससे search accuracy में सुधार होता है।

प्रश्न 5. आधुनिक AI-आधारित सूचना पुनर्प्राप्ति मॉडलों का वर्णन कीजिए।

उत्तर:
आधुनिक AI-आधारित IR models मशीन लर्निंग (Machine Learning), डीप लर्निंग (Deep Learning) तथा Neural Networks पर आधारित होते हैं। ये मॉडल semantic understanding, user intent analysis तथा personalized ranking प्रदान करते हैं। Google, Semantic Scholar और recommendation systems में इनका व्यापक उपयोग किया जा रहा है।

14.18: B) लघु प्रश्न – उत्तर सहित

1. **Vector Space Model क्या है?**
   → यह मॉडल documents एवं query को vectors के रूप में प्रस्तुत करता है।
2. **Cosine Similarity का अर्थ क्या है?**
   → यह query और document के बीच कोण (angle) के आधार पर समानता मापने की विधि है।
3. **Partial matching से क्या तात्पर्य है?**
   → जब document आंशिक रूप से query से मेल खाता हो।
4. **Probability of relevance क्या है?**
   → किसी document के प्रासंगिक होने की संभावना।
5. **Relevance feedback क्या है?**
   → उपयोगकर्ता प्रतिक्रिया के आधार पर search results सुधारने की प्रक्रिया।

C) MCQ:

1. Vector Space Model में documents को किस रूप में represent किया जाता है?
   a) Tables b) Vectors ☑ c) Graph only d) Rules
2. Cosine Similarity किस पर आधारित है?
   a) Distance b) Angle ☑ c) Sorting d) Hashing
3. VSM का मुख्य लाभ क्या है?
   a) Exact match b) Partial match ☑ c) No indexing d) No query

4. VSM में term weights के लिए सामान्यतः क्या प्रयोग होता है?
   a) MARC b) TF–IDF ☑ c) ISBN d) DOI
5. Probabilistic model का focus किस पर है?
   a) Storage b) Probability of relevance ☑ c) Cataloguing d) Circulation
6. Ranking किस model में naturally आती है?
   a) Boolean b) VSM ☑ c) Both Boolean & VSM d) None
7. Relevance feedback का सम्बन्ध किससे है?
   a) Binding b) Query refinement ☑ c) Shelving d) Accession
8. Boolean model के results कैसे होते हैं?
   a) Ranked b) Binary ☑ c) Weighted d) Semantic
9. Probabilistic model का उपयोग कहाँ अधिक होता है?
   a) पुराने card catalogue b) Modern search engines ☑ c) Manual indexing d) Book shelving
10. VSM में query और document similarity अधिक होने का अर्थ है—
    a) Document irrelevant b) Document more relevant ☑ c) No output d) Error
11. Vector Space Model में similarity अधिक होने का अर्थ है—
    a) अधिक distance b) अधिक angle c) अधिक relevance ☑ d) no retrieval
12. TF–IDF का मुख्य उद्देश्य क्या है?
    a) Cataloguing b) Term importance मापना ☑ c) Classification d) Storage
13. Probabilistic model किस प्रकार के search में अधिक उपयोगी है?
    a) Manual search b) Ranked retrieval ☑ c) Shelf browsing d) Accessioning
14. Relevance feedback का मुख्य लाभ क्या है?
    a) Storage बढ़ाना b) Query refinement ☑ c) Index removal d) No effect
15. Neural Network आधारित IR models किस पर कार्य करते हैं?
    a) Rules only b) Logical operators c) Learning from data ☑ d) Card catalogue

14.19 संदर्भ ग्रंथ सूची (References)

1. उत्तराखंड मुक्त विश्वविद्यालय (UOU). **MLIS पाठ्य सामग्री: सूचना पुनर्प्राप्ति प्रणाली** (SLM Modules).
2. कुमार, के. (Kumar, K.). **पुस्तकालय एवं सूचना विज्ञान: मूल सिद्धांत**. (विभिन्न संस्करण).
3. शर्मा, जे. / भारतीय लेखक. **डिजिटल पुस्तकालय एवं सूचना पुनर्प्राप्ति** (संकलित संदर्भ/मॉड्यूल).
4. भारतीय राष्ट्रीय डिजिटल पुस्तकालय/ई-पाठ्य सामग्री: **सूचना खोज एवं पुनर्प्राप्ति (IR) परिचयात्मक नोट्स**.
5. यूजीसी/स्वयंप्रभा/पीठ सामग्री: **सूचना खोज रणनीतियाँ एवं पुनर्प्राप्ति** (प्रासंगिक यूनिट्स).
6. राव, वी. (अनुवाद/हिंदी नोट्स). **सूचना पुनर्प्राप्ति—मॉडल एवं मापन** (शिक्षण नोट्स).
7. राष्ट्रीय पुस्तकालय / LIS प्रशिक्षण सामग्री: **OPAC, डेटाबेस खोज एवं IR आधार**.
8. एन.पी.टी.ई.एल/स्वयं. **Information Retrieval / Search** (हिंदी/द्विभाषी नोट्स जहाँ उपलब्ध).
9. Manning, C. D., Raghavan, P., & Schütze, H. **Introduction to Information Retrieval**. Cambridge University Press.
10. Baeza-Yates, R., & Ribeiro-Neto, B. **Modern Information Retrieval**. Addison-Wesley.
11. Salton, G., & McGill, M. J. **Introduction to Modern Information Retrieval**. McGraw-Hill.
12. Chowdhury, G. G. **Introduction to Modern Information Retrieval**. Facet Publishing.
13. Croft, W. B., Metzler, D., & Strohman, T. **Search Engines: Information Retrieval in Practice**. Pearson.
14. Robertson, S., & Zaragoza, H. **The Probabilistic Relevance Framework: BM25 and Beyond** (foundational reading).
15. Lancaster, F. W. **Information Retrieval Systems: Characteristics, Testing and Evaluation**. Wiley.
16. Van Rijsbergen, C. J. **Information Retrieval**. Butterworths.
17. Belew, R. K. **Finding Out About: A Cognitive Perspective on Search Engine Technology**. Cambridge University Press.

18. Sparck Jones, K. **Foundations of Information Retrieval** (classic contributions on weighting and IR theory).

उत्तराखण्ड मुक्त विश्वविद्यालय
तीनपानी बाईपास रोड , ट्रान्सपोर्ट नगर,
हल्द्वानी -२६३१३९

फ़ोन नं0 : 5946 -261122 , 261123
टॉल फ्री नं ०: 18001804025

Fax No.- 05946-264232, E-mail- info@uou.ac.in
http://uou.ac.in