



खोज की तकनीकियां : मूल और उन्नत

17.1 परिचय

कुछ ढूंढने के लिए भली भांति की गई देखने की गतिविधि को खोज कहते हैं। ग्रंथालय एवं सूचना विज्ञान के संदर्भ में खोज से तात्पर्य है वांछित सूचना ढूंढने के लिए अभिलेखों को ध्यानपूर्वक देखना। सूचना प्राप्त करने के लिए पुनः प्राप्ति के साधन, जैसे प्रसूचियाँ, अनुक्रमणी इत्यादि के संबंध में आप पहले ही अध्ययन कर चुके हैं। इस पाठ में आप पुनः प्राप्ति के उद्देश्य से व्यवस्थित सूचना की आवश्यकता और इसे खोजने के विधि के संबंध में सीखेंगे। इलेक्ट्रॉनिक स्रोतों से सूचना पुनः प्राप्ति के लिए खोज तकनीकियों के आधारभूत पक्षों से भी आपको परिचित कराया जायेगा।



17.2 उद्देश्य

इस पाठ को पढ़ने के पश्चात् आप सक्षम होंगे:

- खोज तकनीकियों को परिभाषित करने;
- शब्दकोश में शब्दों के व्यवस्थापन की व्याख्या करने;
- ग्रंथालय की सामग्री की व्यवस्था एवं पुनः प्राप्ति के लिए शब्दकोश, संख्याओं और संख्यात्मक/वर्गीकृत तकनीकियों का प्रयोग करने;
- सर्च इंजन की परिभाषा करने;
- खोज की प्रक्रिया को पहचानना और खोज प्रश्नावली तैयार करने;
- खोज के प्रचालकों (आपरेटर्स) की भूमिका को जानने;
- बूलियन तर्क (लॉजिक) को परिभाषित करने;

- खोज के प्रकारों को समझने; और
- क्षेत्र आधारित और पूर्ण पाठ्य वस्तु की खोज को सोदाहरण परिभाषित, विस्तृत व्याख्या और इनमें अन्तर स्थापित करने में।

17.3 खोज तकनीकियां

खोज तकनीक एक क्रियाविधि है, जिसके माध्यम से कोई भी व्यक्ति सूचना प्रणालियों में से संबंधित सूचना ढूंढ सकता है। सूचना प्रणाली अंतरंग अथवा ऑनलाइन किसी भी प्रकार की हो सकती है। अंतरंग प्रणाली वह है जिसमें सूचना की पुनः प्राप्ति के उद्देश्यों से, सूचना को किसी संगठन के विस्तार-क्षेत्र के भीतर ही भंडारित किया जाता है। ऑनलाइन सूचना प्रणाली वह प्रणाली है जिसमें इलेक्ट्रॉनिक सूचना स्रोतों को किसी दूरस्थ स्थान पर भंडारित किया जाता है और इन तक अभिगम किसी संचार यांत्रिकी के माध्यम से ही किया जा सकता है। अधिकांश ऑनलाइन सूचना प्रणालियाँ वर्ड वाइड वेब (डब्ल्यू डब्ल्यू डब्ल्यू) के साथ सुसंगत होती हैं और इंटरनेट के माध्यम से इन तक पहुँचा जा सकता है। इन हाउस सूचना प्रणाली में सूचना स्रोत मुद्रित और इलेक्ट्रॉनिक, दोनों रूप में हो सकते हैं। अतः भंडारण क्रियाविधि और खोज तकनीकियाँ, दो विभिन्न पक्ष हैं। हम सूचना के भंडारण और पुनः प्राप्ति के दोनों पक्षों पर चर्चा करेंगे।



पाठगत प्रश्न 17.1

1. खोज तकनीक की परिभाषा लिखें।
2. ऑनलाइन सूचना प्रणाली क्या है?

17.4 भंडारण क्रियातंत्र

संगठन में अंतरंग सूचना प्रणालियों और ऑनलाइन सूचना प्रणालियों का विनिर्माण किसी विशिष्ट वर्ण्य विषय अथवा विषय से संबंधित विशिष्ट सूचना के भंडारण के लिए किया जाता है। इन प्रणालियों में विशिष्ट सूचना ढूंढने के लिए अपना निजी खोज क्रियातंत्र और मार्गनिर्देशों का एक समुच्चय उपलब्ध होता है। ग्रंथालयों और सूचना केन्द्रों में, सूचना मुद्रित और इलेक्ट्रॉनिक, दोनों रूप में उपलब्ध होती है। नीचे, सूचना प्राप्ति में भंडारण की कुछ क्रियातंत्रों और उनकी भूमिका के संबंध में उल्लेख किया गया है:

- (1) शब्दकोश व्यवस्थापन
- (2) संख्यात्मक (न्यूमेरिक) व्यवस्थापन
- (3) वर्गीकृत व्यवस्थापन





टिप्पणी

17.4.1 शब्दकोश व्यवस्थापन

शब्दकोश व्यवस्थापन से तात्पर्य है ऐसी व्यवस्था जिसमें शब्दों को भाषा के वर्णानुक्रमानुसार व्यवस्थित किया जाता है। वर्णानुक्रमानुसार क्रम वह क्रमावली है जो कि भाषा की लिपि में विशिष्ट वर्ण की स्थिति पर आधारित होता है। उदाहरण के लिए अंग्रेजी भाषा में रोमन वर्णों का प्रयोग होता है और इनका क्रम होता है ए, बी, सी, डी,जेड। यहाँ वर्ण “ए” प्रथम स्थान पर है, “बी” दूसरे स्थान पर और इसी प्रकार “जेड” वर्ण इस क्रमावली में छब्बीसवें स्थान पर आता है। इसलिए, शब्दकोश व्यवस्थापन में शब्दों को वर्णों की स्थिति की क्रमावली के अनुसार व्यवस्थित किया जाता है। उदाहरण के लिए:

Action

Ante

Apple

Art

Catalogue

Classification

Search

यहाँ पहले चार शब्द “ए” वर्ण से शुरू हुये हैं और उनकी स्थितियाँ दूसरे, तीसरे अथवा चौथे वर्ण के अनुसार निर्धारित की गई है। इसके उपरान्त दो शब्द “C” वर्ण से शुरू हुये हैं। अतः “C” वर्ण से शुरू होने वाले शब्दों के समूह को “A” शब्द वर्ण से शुरू होने वाले शब्दों के बाद की स्थिति में रखा गया है। इसी प्रक्रिया का अनुसरण करते हुये, इस प्रकार के व्यवस्थापन में शब्दों को व्यवस्थित किया गया है। व्यवस्थापन की इस क्रियातंत्र का अनुसरण प्रसूचियों में उन प्रविष्टियों की व्यवस्था के लिए किया जाता है, जिनमें अभिगम बिन्दु शब्दों के रूप में हैं। उदाहरण के लिए लेखक, शीर्षक, विषय इत्यादि।

17.4.2 संख्यात्मक व्यवस्थापन

संख्यात्मक व्यवस्थापन, वह व्यवस्थापन है जिसमें संख्याओं को आरोही अथवा अवरोही क्रम में व्यवस्थित किया जाता है। उदाहरण के लिए;

123.45

234.15

234.51

435.21

541.23



यहाँ आप देखेंगे कि सभी संख्याओं में पाँच अंको का एक समान समूह, अर्थात् 1 से 5 तक के अंक शामिल है परन्तु उनके सांख्यिकीय मूल्य के अनुसार इनको आरोही क्रम में व्यवस्थित किया गया है और इसी के अनुरूप क्रमावली बनाई गई है। उन ग्रंथालयों में जिनमें ड्यूई दशमलव वर्गीकरण प्रणाली अपनाई गई है, आप पायेंगे कि शैल्फ पर पुस्तकों को सांख्यिकीय क्रम में ही व्यवस्थित किया गया है।

17.4.3 वर्गीकृत व्यवस्थापन

अधिकांश ग्रंथालय, शैल्फ पर पुस्तक की व्यवस्था पुस्तक की क्रामक संख्या (कॉल नंबर) के अनुसार करते हैं। यह कॉल नंबर, वर्ग संख्या, पुस्तक संख्या और संग्रह संख्या का संयुक्त रूप होता है। यह तीनों संख्याएँ, ग्रंथालय द्वारा अपनाई गई प्रणाली के अनुसार, सांख्यिकीय अथवा वर्णसांख्यिकीय (अल्फान्यूमेरिक), कुछ भी हो सकती हैं। अतः जब हमें सांख्यिकीय, वर्ण सांख्यिकीय अथवा वर्गीकृत व्यवस्था की समझ होती है तो शैल्फ से पुस्तकें प्राप्त करना सरल हो जाता है। उदाहरण स्वरूप डीडीसी प्रणाली पर आधारित कुछ क्रामक संख्याओं की व्यवस्था, इस प्रकार होगी।

321.4 RAM

370.1150954 DEM

370.1523 DES

371 ILL 371.3078 KEM

371.32 NIS

371.397 GRE

371.926 BRA

सूचना पुनः प्राप्ति के लिए इन व्यवस्थापनों का अनुसरण करने वाला एक अन्य उदाहरण पुस्तक अनुक्रमणिका से लिया गया है। आपने देखा होगा कि लगभग सभी पुस्तकों के अंत में एक अनुक्रमणिका है। अनुक्रमणिका शब्दों/पदों की एक सूची होती है, जिसके साथ वह पृष्ठ संख्या भी दी जाती है जिस पर वर्ण्य वस्तु में वे शब्द/पद आते हैं। पुस्तक के आकार और प्रकृति के आधार पर इन पदों को पुस्तक अनुक्रमणी में शब्दकोश अथवा वर्गीकृत क्रमानुसार व्यवस्थित किया जाता है। इन व्यवस्थाओं को समझने के उपरान्त आप किसी पुस्तक में से किसी वर्ण्य विषय से संबंधित सूचना सरलता से ढूँढ सकते हैं।



पाठगत प्रश्न 17.2

1. शब्दकोश, संख्यात्मक और वर्गीकृत व्यवस्थापन की परिभाषा लिखें।



टिप्पणी

17.5 सर्च इंजन

इलेक्ट्रॉनिक अथवा डिजिटल माध्यम (मीडिया) से सूचना की खोज, मुद्रित माध्यम में खोज से नितान्त भिन्न है। सूचना को जब इलेक्ट्रॉनिक अथवा डिजिटल रूप में भंडारित किया जाता है तो उस प्रणाली में से संबंधित सूचना ढूँढने के लिए प्रयोक्ता को एक इंटरफेस उपलब्ध कराया जाता है। यह प्रयोक्ता-इंटरफेस एक सॉफ्टवेयर होता है, जिसमें खोज करने के लिए अपेक्षित सूचना का प्रतिनिधित्व करने वाले मुख्य शब्दों (की वर्ड्स) अथवा पदों को स्वीकार करने का प्रावधान होता है। इसमें खोज का परिणाम उसी फार्मेट में सामने आता है जैसा कि उस सॉफ्टवेयर में परिभाषित किया गया होता है। सूचना प्रणाली में से सूचना खोजने के लिए निर्धारित सॉफ्टवेयर को सर्च इंजन कहते हैं। सर्च इंजन को हम इस प्रकार परिभाषित कर सकते हैं कि, “यह एक ऐसा सॉफ्टवेयर है जो इलेक्ट्रॉनिक अथवा डिजिटल सूचना क्षेत्र में से सूचना प्राप्ति के लिए प्रयोग किया जाता है किसी खोजकर्ता द्वारा दिये गये इनपुट के आधार पर उस प्रयोक्ता की पसंद के अनुरूप परिणाम प्रदर्शित करता है। सर्च इंजन में प्रविष्ट किए गये इनपुट को सर्च स्ट्रिंग अथवा प्रश्नावली कहते हैं। यह प्रश्नावली खोजी जाने वाली अपेक्षित सूचना का प्रतिनिधित्व करने वाला एकल पद अथवा पदों का समूह हो सकता है। सर्च इंजन, अपेक्षित प्रश्नावली के आधार पर सूचना की खोज करता है और उस प्रश्नावली के मेल खाते हुये स्रोतों की सूची उपलब्ध करवाता है। यह सूची उसी प्रारूप में प्रदर्शित होगी, जैसा कि सर्च इंजन में इसे तैयार किया गया होगा। सर्च इंजन की प्रकृति के आधार पर, इस सूची में सूचना स्रोतों का संक्षिप्त विवरण समाविष्ट हो सकता है, और इसी के आधार पर खोजकर्ता अपनी इच्छानुसार पूर्ण अभिलेख (रिकॉर्ड) को प्राप्त कर सकता है अथवा उसका संदर्भ दे सकता है अथवा इसके लिए मना भी कर सकता है। आपने इंटरनेट पर अपने ग्रंथालय के ऑनलाइन पब्लिक एक्सेस कैटालॉग ओपैक अथवा लाइब्रेरी ऑफ कांग्रेस ऑनलाइन कैटालॉग (एल सी ओ सी) अथवा पबमैड के साथ-साथ गूगल अथवा याहू पर खोज की होगी यह सभी सर्च इंजन हैं।



पाठगत प्रश्न 17.3

1. किसी सर्च इंजन का विस्तार से वर्णन करें।

17.6 खोज प्रक्रिया

खोज प्रक्रिया कुछ कार्यों का एक समूह (सेट) है जिन्हें अपेक्षित सूचना की प्रभावी ढंग से खोज के लिए निष्पादित किया जाता है। इस खोज के दौरान खोज को क्रियान्वित करने और वाँछित परिणाम प्राप्त करने के लिए कुछ आधारभूत चरणों का अनुपालन करना होता है। ये चरण निम्नलिखित हैं:

- (1) आवश्यकता की पहचान और उसका वर्णन
- (2) खोज रणनीति का विकास
- (3) खोज रणनीति का कार्यान्वयन



- (4) खोज परिणामों की पुनरीक्षा
- (5) खोज परिणामों का सम्पादन
- (6) मूल्यांकन और प्रतिपुष्टि (फीडबैक)

17.6.1 आवश्यकता की पहचान और उसका वर्णन

किसी सूचना व्यवसायी अथवा खोजकर्ता के लिए खोज की अनिवार्यता और इसके उद्देश्य को समझने का अत्यंत महत्त्व है किसी भी विषय (टॉपिक) से संबंधित सूचना की आवश्यकता सामान्य ज्ञान, शोध (रिसर्च) और विकास अथवा किसी भी अन्य उद्देश्य के लिए हो सकती है। खोज की आवश्यकता और इसके उद्देश्य को समझने के बाद, प्रश्नावली का विवरण तैयार किया जाना चाहिए।

सूचना की खोज के इच्छुक और खोजकर्ता में खोज की आवश्यकताओं के संबंध में परस्पर सहमति होनी चाहिए। इस प्रकार की सहमति से संबद्ध और प्रभावी परिणाम प्राप्त करने के लिए खोज की प्रभावी रणनीति बनाने का मार्ग निर्दिष्ट होता है।

17.6.2 खोज रणनीति का विकास

खोज रणनीति में शामिल हैं—प्रश्नावली की अवधारणा की रचना, अवधारणा की विरचना का मुख्य शब्द की भाषा में अनुवाद, विवरणात्मक अथवा पक्ष परिचायक (फेसेट्स), पर्यायों और सहयोजित पदों की पहचान, इत्यादि। रंगनाथन द्वारा प्रदत्त पक्ष (फेसेट्स) की अवधारणा के विश्लेषण (पी एम ई एसटी) के साथ-साथ विशिष्ट विषय की अवधारणा को भी प्रश्नावली की रूपरेखा तैयार करने के लिए प्रभावी साधन के रूप में प्रयुक्त किया जा सकता है। इसके पश्चात् आवश्यकताओं के आधार पर, खोज किये जाने वाले सूचना क्षेत्र (डोमेन) के चयन का भी बहुत महत्त्व है, जैसे कि ग्रंथालय का ओपैक, डाटाबेस और इसी प्रकार के अन्य साधन।

विवरणात्मक पदों की सर्च स्ट्रिंग अथवा प्रश्नावली, उन पदों, शब्दों मुख्य अथवा विवरणात्मक पदों की संयोजन शृंखला होते हैं जो सूचना का प्रतिनिधित्व करते हैं। चूंकि सर्च स्ट्रिंग में शब्दावली का समावेश होता है इसलिए भाषाई विशेषताओं और सूचना की खोज और पुनः प्राप्ति में इनके निहितार्थ का विश्लेषण करना आवश्यक होता है। यहाँ तीन पक्षों, अर्थात् वाक्य विन्यास संबंधी महत्त्व, अर्थ विज्ञान संबंधी महत्त्व और बूलियन संचालकों को समझने की आवश्यकता है।

(क) वाक्य विन्यास संबंधी महत्त्व

सर्च स्ट्रिंग के वाक्य विन्यास के महत्त्व का संबंध उन सूत्रों के प्रकार अथवा संयोजक चिन्हों के साथ होता है। जिसके जरिये अपेक्षित खोज की अवधारणा का प्रतिनिधित्व करने के लिए सर्च इंजन द्वारा मुख्य शब्द अथवा पदों को जोड़ा जाता है। एक 'क' उदाहरण के द्वारा हम किसी प्रश्नावली के वाक्य विन्यास के महत्त्व को समझने का प्रयास करेंगे। दो पद हैं जैसे "Poetry" और "Indians" जिन्हें दो भिन्न संयोजकों, "among" और "by" से जोड़ा गया है। प्रत्येक पद भिन्न अर्थ व्यक्त करता है जो निम्न प्रकार से है:



टिप्पणी

(क) 'Poetry among Indians' का अर्थ है 'What is the status of Poetry among Indian's अर्थात् भारतीयों में कविता का क्या स्तर है अथवा What is the approach of Indians towards Poetry?' अर्थात् 'भारतीयों का कविता के प्रति क्या दृष्टिकोण है?'

(ख) 'Poetry by Indians' से तात्पर्य है भारतीयों द्वारा सृजित कविता।

(ख) शब्दार्थपरक महत्व

किसी सर्च स्ट्रिंग के शब्दार्थपरक महत्व का संबंध स्ट्रिंग के अर्थ में इस बात से है कि सर्च इंजन द्वारा अपेक्षित सूचना और अर्थ प्रदर्शन का संदर्भ क्या है? संभावित खोज की अवधारणा के अर्थ को निश्चित करने और सर्च इंजनों द्वारा उन्हें समझे जाने के लिए हम प्रचालकों का प्रयोग, सर्च इंजनों द्वारा स्वीकार्य मुख्य शब्द के संयोजकों के रूप में करते हैं। किसी प्रश्नावली की अर्थमूलक महत्ता को हम नीचे दिये गये दो उदाहरणों के माध्यम से समझ सकते हैं:

(1) 'Contribution of Indian Society in Mathematics' प्रश्न (क्वेटी) का अर्थ है गणित के क्षेत्र में भारतीय समाज का योगदान.

(2) "Contribution of Mathematics in Indian Society" प्रश्नावली का अर्थ है भारतीय समाज का निर्माण करने में गणित का योगदान। ये दोनों उदाहरण किसी प्रश्न के अर्थपरक महत्व के संबंध से हमें स्पष्ट अर्थ का बोध करवाते हैं। मुख्य शब्द और संयोजक शब्द का समुच्चय पद, विभिन्न क्रम में लिखे जाने पर विभिन्नता प्रदान करेगा।

(ग) बूलियन संचालक

बूलियन संचालक बहुत ही सरल शब्द होते हैं (AND, OR and NOT) जिन्हें खोज क्रिया में मुख्य शब्द को जोड़ने अथवा विलग करने के लिए संयोजकों के रूप में प्रयुक्त किया जाता है। इनका प्रयोग खोज पदों में परस्पर संबंधों को जोड़ने अथवा परिभाषित करने के लिए किया जाता है। अतः इस प्रकार परिणामों को और अधिक सटीक और उपयोगी बनाया जाता है। इन तीनों पदों को, सर्च इंजनों के निर्माताओं की व्यापक स्वीकृति प्राप्त है। सूचना की खोज में संचालकों के रूप में प्रयोग किए जाने के दौरान इनका अर्थ पूर्णतया सुस्पष्ट होता है। बूलियन तर्क के तीन संचालक हैं तार्किक योग (+) OR, तार्किक उत्पाद (प्रॉडक्ट) (x) AND, और तार्किक अन्तर (-) Not। सूचना पुनः प्राप्ति की सभी प्रणालियों में उपयोक्ताओं को अपने प्रश्नों को इन संचालकों का प्रयोग करके अपनी अभिव्यक्ति करने की स्वीकृति प्राप्त है। आइये अब हम इन तीनों संचालकों की संलिप्तता के बारे में समझते हैं।

OR Operator : OR ऑपरेटर, खोजकर्ताओं को खोज के पदों में विकल्प विनिर्दिष्ट करने की अनुमति देते हैं। जब भी 'OR ऑपरेटर' का प्रयोग करके, सर्च स्ट्रिंग बनाई जाती है तो सर्च इंजन उन सभी स्रोतों की पुनः प्राप्ति करता है जिनमें 'OR' से संबद्ध कोई भी पद अथवा मुख्य शब्द मौजूद होते हैं। उदाहरण के लिए यदि हम "Student OR education" जैसी सर्च स्ट्रिंग बनाते हैं और इसे खोजते हैं तो खोज के परिणाम के रूप में हमें उस प्रणाली में उपलब्ध उन सभी स्रोतों के संदर्भों की सूची प्राप्त होगी जिनमें student (विद्यार्थी) अथवा education (शिक्षा) शब्द मौजूद होंगे।



टिप्पणी

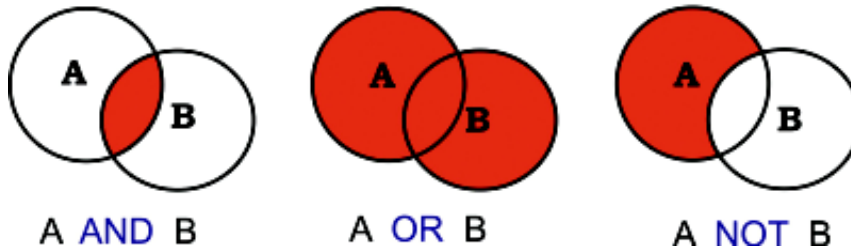
AND Operator : AND ऑपरेटर का प्रयोग दो अथवा अधिक पदों को जोड़ने के लिए किया जाता है। जब भी AND Operator का प्रयोग करके सर्च स्ट्रिंग बनाई जाती है, तो सर्च इंजन उन सभी स्रोतों की पुनः प्राप्ति करता है जहाँ 'AND' शब्द से जुड़े कोई भी पद अथवा मुख्य शब्द मौजूद होंगे। उदाहरण के लिए यदि हम "Student And education" जैसा कोई संपर्क सूत्र बनाते हैं और उसे खोजते हैं तो खोज के परिणाम में हमें उन सभी संसाधनों की सूची प्राप्त होगी जिनमें Student and education दोनों पद मौजूद होंगे।

NOT Operator : NOT ऑपरेटर का प्रयोग स्रोतों के समूह में से किसी पद को विलग करने के लिए किया जाता है। उदाहरण के लिए यदि हम "Student Not Education" का खोज सूत्र बनाते हैं और इसे खोजते हैं तो खोज के परिणामस्वरूप हमें, उस प्रणाली में उपलब्ध उन सभी संदर्भों की सूची प्राप्त होगी जिनमें student पद तो होगा परन्तु 'education' मौजूद नहीं होगा।

नीचे दिये गये रेखा चित्रों से आप बूलियन ऑपरेटर्स की संलिप्तता को समझ सकते हैं। यहाँ 'A' Students (विद्यार्थियों) के समूह को दर्शाता है और 'B' education (शिक्षा) के समूह को चित्र 17.1 दर्शाता है।

बूलियन ऑपरेटर्स

BOOLEAN OPERATORS



चित्र 17.1 बूलियन की संलिप्तता



पाठगत प्रश्न 17.4

1. खोज रणनीति की परिभाषा लिखें।

17.6.3 खोज रणनीति का कार्यान्वयन

खोज को कार्यान्वित करने से पहले, खोजकर्ता को डाटा की उस रूपरेखा का ज्ञान होना चाहिये जिसे उस सूचना प्रणाली के द्वारा उस डाटा के भंडारण के लिए अपनाया गया होगा। प्रणाली आधारित सर्च इंजनों को किसी डाटाबेस में से सूचना खोजने के लिए उसी प्रकार निर्मित किया जाता है जिस प्रकार इसके रचयिताओं ने इसे तैयार किया होगा। जैसे कि OPAC में, हम



टिप्पणी

'Tagore रबीन्द्रनाथ' के रूप में अपनी प्रश्नावली डालते हैं और लेखक के नाम के क्षेत्र में इसे खोजते हैं तो डाटा बेस में से केवल वही अभिलेख पुनः प्राप्त और प्रदर्शित होंगे जिन्हें इस लेखक द्वारा सृजित किया गया होगा। परन्तु, यदि हम इसी प्रश्नावली को शीर्षक के क्षेत्र में खोजने के लिए निर्दिष्ट करेंगे तो वे अभिलेख प्रदर्शित होंगे जिनमें 'टैगोर' रबीन्द्रनाथ का नाम 'शीर्षक' में अथवा शीर्षक के किसी भाग में मौजूद होगा। इसका तात्पर्य यह है कि 'टैगोर, रबीन्द्रनाथ' के संबंध में लिखी गई संपूर्ण सामग्री के संदर्भों की सूची खोज के परिणामस्वरूप प्राप्त होगी। खोज की आवश्यकता और उद्देश्य और खोजकर्ता के अनुभव के आधार पर सर्च इंजनों की विशेषताओं का प्रयोग करते हुये खोज रणनीति को कार्यान्वित किया जा सकता है। अतः प्रभावी परिणाम प्राप्त करने के लिए खोजकर्ता को खोज की किस्मों और उसकी सलिप्तता के संबंध में जानकारी होनी चाहिए। खोज के विभिन्न प्रकार हैं यथा:

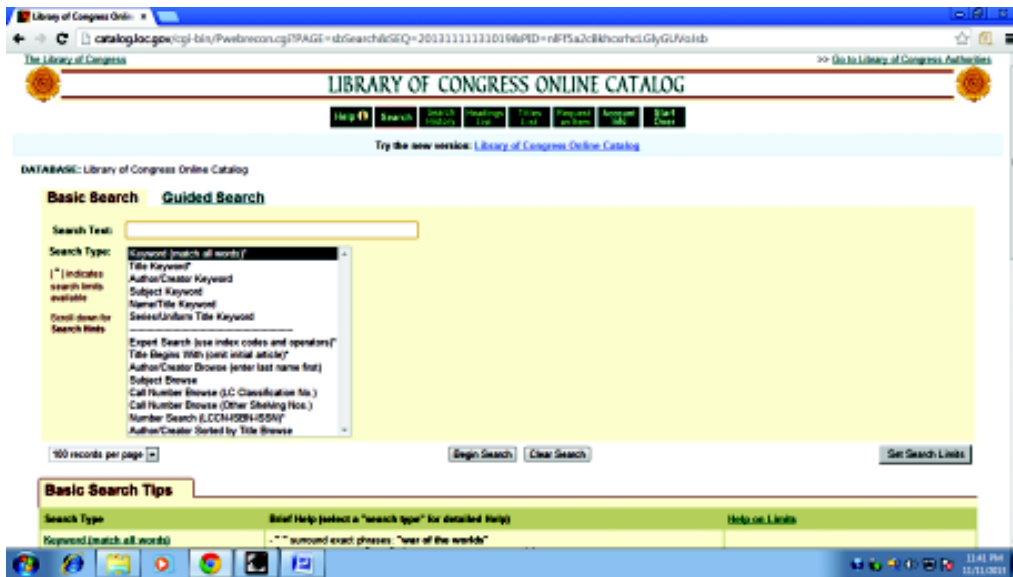
- (क) क्षेत्र आधारित खोज (फील्ड बेस्ड सर्च)
- (ख) संपूर्ण पाठ्य वस्तु की खोज (फुल टेक्स्ट खोज)
- (ग) संक्षेपन करके खोज (ट्रंकेशन सर्च)
- (घ) सान्निध्यता खोज (प्रॉक्सिमिटी सर्च)
- (ङ.) सीमितकरण खोज (लिमिटिंग सर्च)
- (च) दायराबद्ध खोज (रेंज सर्च)
- (छ) सरल खोज (सिम्पल सर्च)
- (ज) उन्नत खोज (एडवांस्ड सर्च)

(क) क्षेत्र आधारित खोज (फील्ड बेस्ड सर्च)

अपेक्षित सूचना प्राप्त करने के लिए किसी विशिष्ट क्षेत्र के डाटाबेस पर की गई खोज को क्षेत्र आधारित खोज कहते हैं। जैसा कि आप जानते हैं, प्रसूची की संपूर्ण सूचना को संदर्भ ग्रंथसूची परक डाटाबेस में विविध क्षेत्रों में भंडारित किया जाता है। यदि आप लेखक के संबंध में खोज करना चाहते हैं, तो सर्च इंजन को लेखक क्षेत्र के लिए निर्देश दें अथवा यदि आप 'शीर्षक' अथवा 'विषय' के संबंध माध्यम से खोज करना चाहते हैं, तो आपको सर्च इंजन को शीर्षक या विषय क्षेत्र का निर्देश दें यदि 'सेन, अमर्त्य' को लेखक क्षेत्र में खोजा जाये, तो खोज के परिणामस्वरूप आपको इस लेखक द्वारा रचित सभी रचनाएँ और सृजित कार्य प्रदर्शित किया जायेगा। जबकि यही खोज जब शीर्षक क्षेत्र में की जायेगी तो इस लेखक से संबद्ध सृजित की गई रचनाएँ प्रदर्शित होंगी। उदाहरण के लिए लाइब्रेरी ऑफ कांग्रेस ऑनलाइन कैटलॉग (एल ओ सी) में 'सेन, अमर्त्य' के रूप खोज सूत्र के द्वारा खोजे जाने वाले क्षेत्र के रूप में 'ऑथर/क्रियेटर' का चयन किया गया। जब 'ऑथर/क्रियेटर' की खोज की गई तो सर्च इंजन द्वारा दिये गये परिणाम में 7 प्रलेखों की सूची दर्शाई गई, और जब शीर्ष क्षेत्र को खोजा गया तो इसने 157 प्रलेखों की सूची सामने आई। खोज और परिणाम को नीचे चित्र 17.2 से 17.5 में दर्शाया गया है।



टिप्पणी



चित्र 17.2 एल सी ओ सी पर खोज



चित्र 17.3 एल सी ओ सी पर खोज जहाँ अर्मत्य सेन को ऑथर/क्रिएटर के क्षेत्र में खोजा गया।

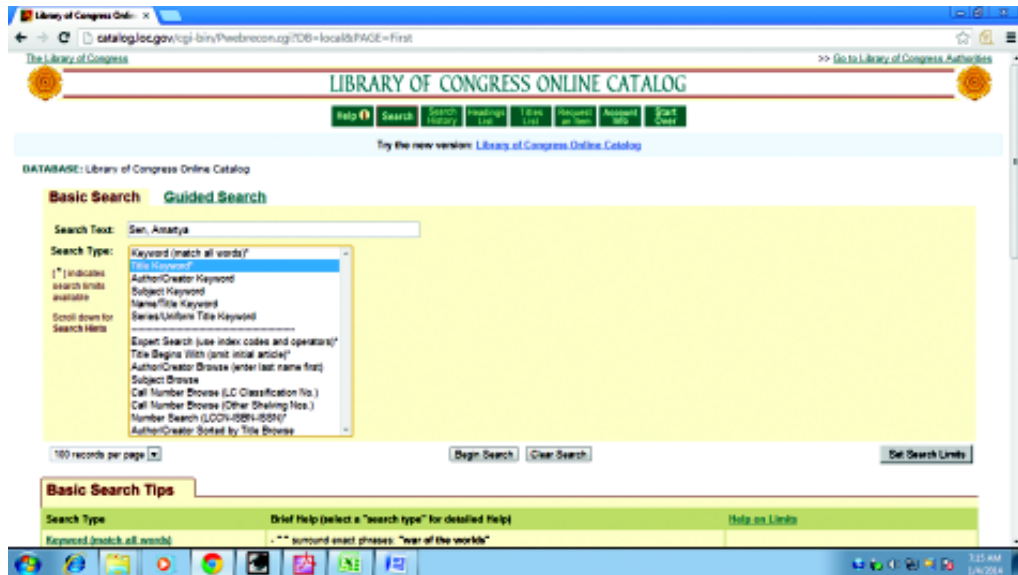
मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



टिप्पणी

खोज की तकनीकियां : मूल और उन्नत



चित्र 17.4 एलसीओ पर खोज



चित्र 17.5 एलसीओसी पर खोज, जहां सेन अर्मत्य को शीर्षक (टाइटल) क्षेत्र में खोजा गया।

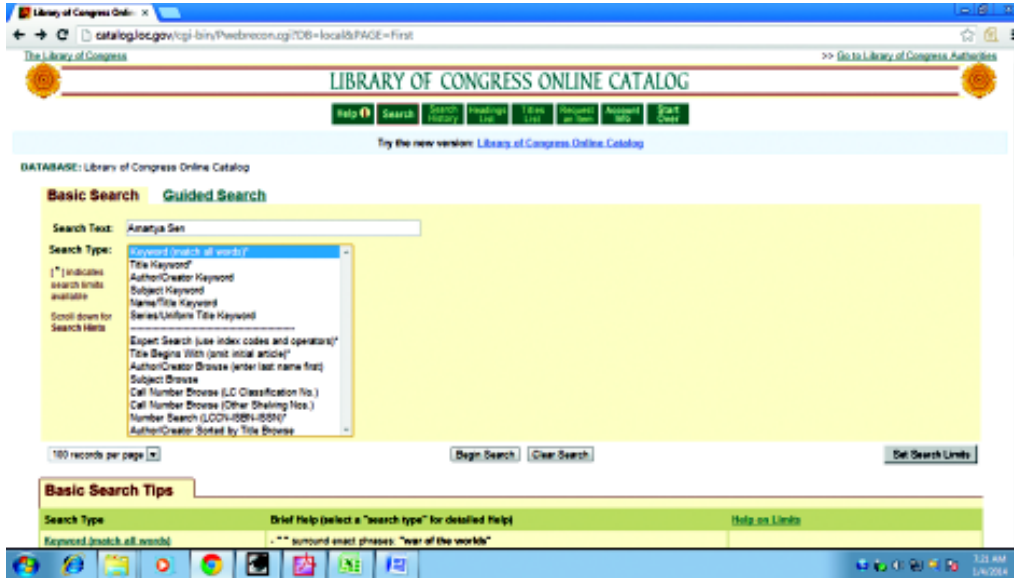
(ख) पूर्ण पाठ्य वस्तु (टैक्स्ट) की खोज

पूर्ण पाठ्य वस्तु की खोज, खोज की एक ऐसी क्रियाविधि है, जो डाटाबेस के प्रत्येक क्षेत्र में खोज करती है और उन सभी अभिलेखों को निकाल लेती है जो प्रश्नावली से मेल खाते हैं। उदाहरण के लिए जब वही खोज (अर्मत्य सेन) मुख्य शब्द के विकल्प के साथ LCOC पर



टिप्पणी

की गई, जो पूर्ण पाठ्य वस्तु की खोज के रूप में कार्य करता है, तो 193 अभिलेखों की सूची प्राप्त हुई। इससे यह प्रदर्शित होता है कि पूर्ण पाठ्य की खोज में सफलता प्राप्ति के प्रयासों (हिट्स) में वृद्धि हुई क्योंकि इसके परिणामस्वरूप वे सभी परिणाम सामने आए जिनमें किसी भी क्षेत्र में सभी अभिलेखों में किसी भी रूप में 'सेन, अमर्त्य' मौजूद था। इस खोज परिणाम को चित्र 17.6 और चित्र 17.7 में दर्शाया गया है।



चित्र 17.6 : LCOC पर खोज



चित्र 17.7 : LCOC पर खोज, जहाँ Sen, Amartsen को मुख्य शब्द मिलान के रूप में खोजा गया।



टिप्पणी

(ग) संक्षेपन खोज

संक्षेपन खोज, खोज की वह तकनीक है जिसमें किसी ऐसे शब्द के विभिन्न रूपों के लिए खोज क्रियान्वित की जाती है जिनका विकास एक सामान्य मूल धातु-पद से हुआ हो। सूचना पुनः प्राप्ति प्रणाली में यह तरीका सर्वाधिक व्यापक रूप से अपनाया जाता है। इस तकनीक में मूल धातु-पद को संक्षेपन चिन्ह के साथ खोज क्रियान्वित की जाती है। उदाहरण के लिए यदि हम 'India' की खोज करते हैं तो वे सभी अभिलेख पुनः प्राप्त होंगे जिनमें 'India' पद से जुड़ा पूर्ण या आंशिक रूप से कोई शब्द मौजूद होगा। परिणाम में आए अभिलेखों की सूची में इन प्रलेखों के प्रभाव-क्षेत्र (डोमेन) में India, Indian, Indiana, Indian nation अथवा इसी प्रकार के शब्दों की सूची प्राप्त होगी।

(घ) सानिध्यता खोज

सानिध्यता खोज, खोज की वह तकनीक है, जो खोजकर्ता को दो पदों के परस्पर अन्तर को परिभाषित करने की अनुमति देती है। यह दो पद, चाहे एक दूसरे के साथ-साथ हों, अथवा, खोज के पदों के बीच में एक अथवा एक से अधिक शब्द आते हों, अथवा खोज के पद एक ही अनुच्छेद में प्रयुक्त हों, फिर चाहे उनके बीच में अन्य शब्द हस्तक्षेप कर रहे हो, इत्यादि। इस उद्देश्य के लिए विभिन्न सर्च इंजन प्रचालकों के विभिन्न समुच्चयों का प्रयोग करते हैं।

(ङ) सीमितिकरण खोज

सीमितिकरण खोज तकनीक में, एक खोजकर्ता स्ट्रिंग को, डाटाबेस की संरचना (आर्किटेक्चर) के अनुसार सीमित कर देता है और इसी स्ट्रिंग के विभिन्न क्षेत्रों में खोजता है। उदाहरण के लिए यदि कोई खोजकर्ता 'Development as freedom by Amartya Sen' को खोज रहा है तो इस स्ट्रिंग को दो छोटे स्ट्रिंग में तोड़ दिया जाएगा, जैसे 'Development as freedom' and 'Amartya Sen'. स्ट्रिंग 'Development as freedom' को टाइटल फील्ड में डाला जाएगा और स्ट्रिंग सब-स्ट्रिंग 'Amartya Sen' को लेखक क्षेत्र में डाला जाएगा। और इसके बाद खोज को क्रियान्वित किया जाएगा।

(च) दायराबद्ध खोज

दायराबद्ध खोज (रेंज सर्च) तकनीक, एक ऐसी तकनीक है जो खोजकर्ताओं को किन्हीं विशिष्ट डाटा प्रभाव-क्षेत्रों में ही अभिलेखों को चुनने की अनुमति देती है। यह तकनीक संख्यात्मक डाटा खोज के लिए अधिक उपयुक्त हैं। एक सर्च इंजन से दूसरे सर्च इंजन में इनके संचालकों के अर्थ में भिन्नता होती है। सामान्यतया प्रयोग में लाये जाने वाले कुछेक संचालक (ऑपरेटर्स) हैं:

- उससे बड़ा (ग्रेटर दैन) (>)
- उससे कम (लैस दैन) (<)
- उसके समान (=)
- उसके असमान (1= अथवा <>)



- उससे बड़ा अथवा उसके समान (ग्रेटर दैन or equal to) ($> =$)
- उससे कम अथवा उसके बराबर (लैस दैन or equal to) ($< =$)

उदाहरण के लिए यदि हम उल्लेख करें कि प्रकाशन वर्ष 2000 $> =$ है तो परिणामस्वरूप उन सभी स्रोतों की सूची प्राप्त होगी जो 2000 AD और उस के बाद प्रकाशित हुये हैं।

(छ) सरल खोज

सरल खोज एक ऐसी तकनीक है जिसमें एक खोजकर्ता, सर्च इंजन के व्यवहार को अथवा डाटाबेस की संरचना को समझे बगैर अथवा संचालकों और संयोजकों के प्रभाव को जाने बिना बहुत ही सरल प्रारूप मुख्य शब्द खोज के लिए प्रयुक्त करते हैं। लगभग सभी सर्च इंजन, सरल खोज के प्रयोग की सुविधा उपलब्ध कराते हैं। सरल खोज फुल टेक्स्ट सर्च के मॉडल के अनुरूप कार्य करती है जिसके संबंध में ऊपर चर्चा की गई है।

(ज) उन्नत खोज (एडवांस्ड सर्च)

उच्च स्तरीय खोज वह खोज है जिसके माध्यम से एक खोजकर्ता, विभिन्न प्रकार के साधनों और क्रियाविधियों का प्रयोग करके खोज करता है, ताकि वह सुस्पष्ट और सुसंबद्ध परिणाम प्राप्त कर सके। इस तकनीक में एक खोजकर्ता सर्च इंजन द्वारा उपलब्ध कराये गये प्रचालकों और परिसीमाओं का प्रयोग करके खोज-सूत्र का निर्माण करता है। ऊपर विचार किए गए विभिन्न तरीकों के संयोजन से, सूचना की खोज करना भी इसी वर्ग में आता है। यहाँ खोज सूत्र के प्रत्येक पद के क्षेत्र (स्कोप) को, सर्च इंजन में उपलब्ध सुविधानुसार परिभाषित किया जा सकता है। उन्नत उच्च स्तरीय खोज के विविध पक्षों पर हम पाठ 18 में विचार करेंगे।



पाठगत प्रश्न 17.5

1. क्षेत्र आधारित (फील्ड बेस्ड) खोज क्या है?
2. पूर्ण पाठ्य-वस्तु (टेक्स्ट) खोज की व्याख्या करें।
3. संक्षेपन खोज (ट्रंकेशन सर्च) की परिभाषा लिखें।
4. दायरे बद्ध खोज के संबंध में चर्चा करें।

17.6.4 पुनरीक्षा खोज परिणाम

खोज परिणामों का सर्वोत्तम पुनरीक्षक होता है प्रयोक्ता परन्तु सूचना व्यावसायिकों को भी सूचना पुनःप्राप्ति प्रणालियों के मूल्यांकन के लिए दिये गये मानकों के आधार पर खोज परिणामों की पुनरीक्षा (समीक्षा) करनी चाहिए।

17.6.5 खोज परिणामों का सम्पादन

खोज परिणामों के सम्पादन में, खोज परिणामों को प्रयोक्ता के लिए उपयोगी प्रारूप (फार्मेट)

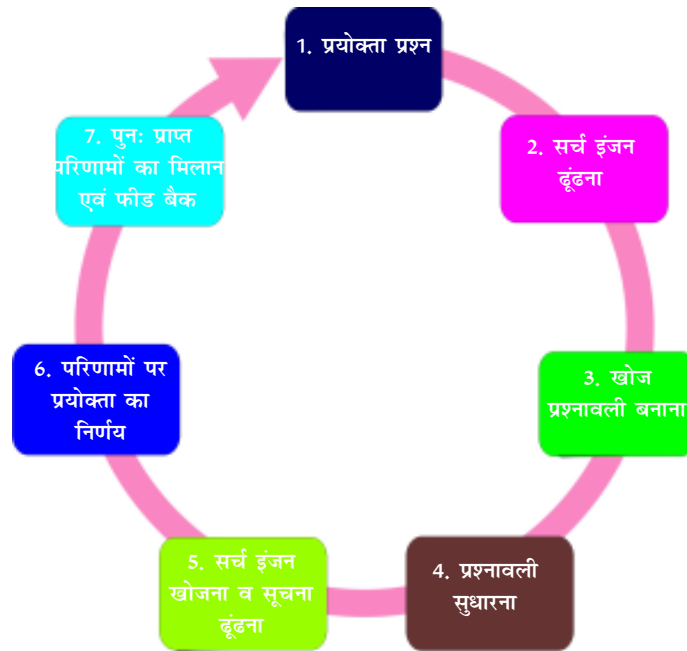


में परिवर्तित करने का कार्य शामिल है। इस कार्य में परिणामों का भली-भांति व्यवस्थित पैकेज में प्रबंधन करना, महत्वपूर्ण तत्त्वों को रेखांकित करना, इन महत्वपूर्ण तत्त्वों में और अधिक सूचना जोड़ना और सूचना को पुनः प्रारूपित करना शामिल है ताकि यह प्रयोक्ताओं की अपेक्षाओं के अनुरूप बन सके।

17.6.6 मूल्यांकन एवं प्रतिपुष्टि (फीडबैक)

खोज परिणामों के मूल्यांकन में प्रयोगकर्ताओं और खोजकताओं दोनों की भागीदारी शामिल होती है। परिणामों की गुणवत्ता और मात्रा का आकलन किया जाता है और यदि आवश्यकता हो, तो इस प्रक्रिया को पुनः परिभाषित और पुनः प्रारंभ किया जा सकता है, यदि प्रयोक्ता अंतिम परिणाम से संतुष्ट न हो।

ऊपर जिस खोज प्रक्रिया की विवेचना की गई है उसके आधार पर नीचे चित्र 17.8 में दर्शाए गये एक सरल मॉडल का प्रावधान किया जा सकता है।



चित्र 17.8 खोज प्रक्रिया का मॉडल



पाठगत प्रश्न 17.6

1. एक खोज परिणाम की पुनरीक्षा/समीक्षा कैसे की जाती है?
2. खोज परिणामों के संपादन की क्या आवश्यकता है?



आपने क्या सीखा

- किसी भी सूचना प्रणाली से सूचना पुनः प्राप्ति के लिए प्रयुक्त मानक क्रियाविधि को, सूचना खोज तकनीक कहते हैं।
- खोज तकनीक वह क्रियाविधि है, जिसके माध्यम से कोई भी व्यक्ति सूचना प्रणालियों में से संबंधित सूचना ढूँढ सकता है। सूचना प्रणाली इन हाउस या ऑनलाइन दोनों प्रकार की हो सकती है।
- भंडारण की क्रियाविधि शब्दकोशीय, सांख्यिकीय और सूचना की वर्गीकृत व्यवस्था हो सकती है।
- खोज प्रक्रिया में कुछ कार्यों के समूह का अनुपालन करना होता है, जैसे:
 - प्रयोक्ता की आवश्यकतानुसार सूचना खोज निर्धारित करना,
 - खोज रणनीति की संरचना;
 - खोजी जाने वाली सूचना प्रणाली का चुनाव और तदानुसार सर्च इंजन की खोज;
 - मुख्य शब्द और संचालकों का प्रयोग करते हुये, खोज प्रश्नावली अथवा खोज सूत्रों का निर्माण करना जो प्रयोगकर्ताओं की आवश्यकताओं के वाक्य विन्यासात्मक मूल्यों को प्रस्तुत करें और खोज इंजन जिस अर्थ को दर्शाए उसके शब्द मूल्य को भी प्रस्तुत कर सकें;
 - खोज क्रियान्वित करना
 - परिणाम का मूल्यांकन यदि आवश्यकता हो तो खोज प्रक्रिया को पुनः छांटना (फिल्टर करना) अथवा पुनः परिभाषित या पुनः प्रारंभ करना; और
 - खोज परिणामों को इस प्रारूप में प्रस्तुत करना कि यह प्रयोक्ता के लिए उपयोगी प्रारूप में आ जाएँ।
- सुसंगत और प्रभावी खोज परिणाम प्राप्त करने के लिए एक खोजकर्ता को खोज के सभी प्रकारों और उन्हें कार्यान्वित करने के कौशल की जानकारी होनी चाहिए।



पाठांत प्रश्न

1. सूचना पुनः प्राप्ति के लिए खोज तकनीकों और उनकी आवश्यकताओं के संबंध में विस्तृत विवरण लिखें।
2. खोज प्रणालियों की सूची बनाएँ और इसके प्रत्येक चरण के संबंध में संक्षिप्त। टिप्पणी दें।
3. एक सर्च स्ट्रिंग के शब्दार्थक मूल्य का विवरण लिखें



टिप्पणी



4. दो मुख्य शब्द 'A' और 'B' को जोड़ने के दौरान, बूलीयन संचालकों और उनके प्रभाव का विस्तृत विवरण दें।
5. एक उन्नत खोज में सरल खोज से कितना अंतर है?



पाठगत प्रश्नों के उत्तर

17.1

1. जिस क्रियाविधि के द्वारा हम किसी सूचना प्रणाली में से संबंधित सूचना ढूँढते हैं उसे खोज तकनीक कहते हैं।
2. ऑनलाइन सूचना प्रणाली एक ऐसी प्रणाली है जहाँ इलेक्ट्रॉनिक स्रोतों का किसी दूरस्थ स्थान पर भंडारण किया जाता है और किसी संचार क्रियाविधि (कम्प्यूनिकेशन मकैनजम) के द्वारा ही इन तक पहुँचा जा सकता है।

17.2

1. शब्दकोशीय व्यवस्थापन, एक ऐसी व्यवस्था है जिसमें शब्दों को संबद्ध भाषा के वर्णानुक्रमानुसार व्यवस्थित किया जाता है। संख्यात्मक व्यवस्था में, संख्याओं को आरोही अथवा अवरोही क्रम में व्यवस्थित किया जाता है। वर्गीकृत व्यवस्था एक ऐसी व्यवस्था है जिसमें शब्द अथवा संख्याये अथवा दोनों के सम्मिश्रित रूप को किन्ही विशेषताओं के आधार पर पहले समूहीकृत किया जाता है। तत्पश्चात् इन्हें किसी शब्दकोशीय या संख्यात्मक रूप अथवा इन दोनों के सम्मिश्रित के रूप में व्यवस्थित किया जाता है।

17.3

1. कोई भी सर्च इंजन एक सॉफ्टवेयर होता है जो किसी इलेक्ट्रॉनिक या डिजिटल सूचना डोमेन से सूचना खोजने के काम आता है। किसी खोजकर्ता द्वारा अपेक्षित प्रश्नावली के आधार पर सर्च-इंजन प्रयोक्ता के लिए उपयुक्त फॉर्मेट में खोज परिणामों को प्रदर्शित करता है।

17.4

1. खोज रणनीति, प्रश्नावली की अवधारणा की संरचना के विकास, अवधारणा की संरचना को मुख्य शब्द व्याख्याकारक शब्द अथवा विविध पहलुओं की भाषा में अनूदित करने और पर्यायों और सहयोजित पदों की पहचान करने इत्यादि की प्रक्रिया है।

17.5

1. क्षेत्र आधारित खोज वह खोज है जहाँ अपेक्षित सूचना की खोज डाटाबेस के विशिष्ट क्षेत्र में क्रियान्वित की जाती है।
2. पूर्ण वर्ण्य-विषय की खोज, खोज की एक ऐसी क्रियाविधि है, जो डाटाबेस के हरेक क्षेत्र

में खोज को क्रियान्वित करती है और उन सभी अभिलेखों को खोज निकालती है जो प्रश्नावली से मेल खाते हैं।

3. संक्षेपन खोज एक ऐसी तकनीक है, जिसमें खोज किसी ऐसे शब्द के विभिन्न रूपों पर क्रियान्वित की जाती है जिनका जन्म उस एक सामान्य मौलिक शब्द से हुआ होगा।
4. दायराबद्ध खोज तकनीक में एक खोजकर्ता, डाटाबेस के रचनाशिल्प (आर्किटेक्चर) के अनुसार स्ट्रिंग को तोड़ता है और इसी के समान सूत्र के विभिन्न पदों को विभिन्न क्षेत्रों में खोजता है।

17.6

1. खोज परिणाम की पुनरीक्षा/समीक्षा का सर्वोत्तम उपाय है प्रयोक्ता से प्रतिपुष्टि (फीडबैक) लेना। यदि प्रयोगकर्ता संतुष्ट है तो खोज के परिणाम को सफल माना जा सकता है।
2. सर्च-इंजनों द्वारा प्रदर्शित परिणाम, हमेशा प्रयोक्ता के वांछित प्रारूप में नहीं होते। अतः परिणाम को प्रयोगकर्ता द्वारा वांछित फॉर्मेट में लाने के कारण, उसका संपादन करना अनिवार्य हो जाता है।

पारिभाषित शब्दावली

वर्णानुक्रमानुसार व्यवस्था (Alphabetical Arrangement) : ऐसी व्यवस्था, जिसमें पदों को उस विशेष भाषा के वर्णानुक्रम के अनुसार लिखा जाता है।

आरोही (Accending) : बढ़ते हुये क्रम में।

अवरोही क्रम में (Descending Order) : घटते हुये क्रम में।

सूचना प्रभाव-क्षेत्र (Information Domain) : ऐसी सूचना प्रणाली जिसमें सूचना को अथवा इसके स्रोतों को पुनः प्राप्ति के उद्देश्य से तार्किक रूप में संग्रहित किया जाता है।

एलसीओसी (LCOC) : लाइब्रेरी ऑफ कांग्रेस की ऑनलाइन प्रसूची (कैटालॉग)

संख्यात्मक व्यवस्था (Numeric Arrangement) : आरोही अथवा अवरोही क्रम में संख्याओं की व्यवस्था

ओपेक (OPAC) : मुक्त सार्वजनिक अभिगम प्रसूची व्यक्तित्व, पदार्थ, ऊर्जा, स्थान और काल वर्गीकरण का कॉलन Personality Matter, Energy, Space and Time :

पबमेड (PubMed) : एक नैशनल लाइब्रेरी ऑफ मेडिसन के ऑनलाइन डाटाबेस में, MEDLINE से जैव चिकित्सीय साहित्य, जीवन-विज्ञान संबंधी पत्रिकाएँ (लाइफ साइंस जर्नल्स) और ऑनलाइन पुस्तकों के उद्धरण शामिल हैं। इन उद्धरणों में pubmed central और Publisher web sites में से पूर्ण पाठ्य वस्तु के संपर्क सूत्र (links) भी सम्मिलित हो सकते हैं।



मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



टिप्पणी

खोज की तकनीकियां : मूल और उन्नत

सॉफ्टवेयर (Software) : कम्प्यूटर में उपलब्ध कोई भी कार्यक्रम (Programme) जिसे कार्यान्वित किया जा सकता हो जैसे Libsys WINISIS इत्यादि।

यूजर इंटरफेस (User interface) : एक ऐसा फलक (उपकरण) जो प्रयोक्ता और प्रणाली में परस्पर संबंध बनाता है (के बीच काम करता है)

www.World Wide Web: इंटरनेट के माध्यम से परस्पर जुड़े हुये (इंटरलिंकड) हाइपरटेक्स्ट प्रलेखों तक पहुँचने की एक प्रणाली है। Web browser (वेब ब्राउजर) के जरिए कोई भी व्यक्ति उन वेब पृष्ठों को देख सकता है जिनमें पाठ्यवस्तु (टेक्स्ट) वीडियो, चित्र और अन्य मल्टीमीडिया शामिल हो सकता है और हाइपरलिंक के माध्यम से इनमें परस्पर नेविगेट किया जा सकता है

वेबसाइट्स

<http://catalog.loc.gov/help/titles.htm>

<http://www.ncbi.nlm.nih.gov>

<http://www.internettutorials.net/basic-search-techniques.asp>

<http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>