



18

सूचना तकनीक : वेब आधारित खोज

18.1 परिचय

इंटरनेट, विश्व में सूचना का सबसे बड़ा भंडार बन चुका है। इसे हम वैश्विक ग्रंथालय (ग्लोबल लाइब्रेरी) भी मान सकते हैं। जहाँ विभिन्न भाषाओं और प्रारूपों में, डिजिटल रूप में विविध प्रकार की सूचना भंडारित है। वेब पर मौजूद सूचना की मात्रा अत्यन्त विशाल है और इसके आकार का अनुमान लगाना लगभग असंभव हो चुका है। इसके आकार और भंडारण की क्रियाविधि के कारण, संबद्ध और सुसंगत सूचना ढूँढ पाना एक बहुत ही कठिन कार्य बन गया है। इस विशाल भंडार में से सूचना ढूँढने के लिए हम सर्च इंजनों का प्रयोग करते हैं। इंटरनेट पर हजारों सर्च इंजन उपलब्ध है। उदाहरण के लिए यदि आप [http://www. The search engine list.com/](http://www.TheSearchEngineList.com/) पर जायेंगे, तो आपको सर्च इंजनों की एक वर्गीकृत सूची मिलेगी। यह सूची वर्ग के अनुसार (कैटेगरी-वाइज) है और इसमें विविध क्षेत्रों के सर्व-प्रयोज्य (ऑल-पर्पज) सर्च इंजन शामिल हैं जैसे लेखाकरण (अकाउंटिंग), ब्लॉग्स, पुस्तकें, विधिक, चिकित्सीय इत्यादि।

पाठ 17 में आपने खोज तकनीक की बुनियादी अवधारणा के संबंध में पढ़ा। यहां आप वेब पर सूचना ढूँढने के विविध पक्षों के बारे में सीखेंगे।



18.2 उद्देश्य

इस पाठ को पढ़ने के पश्चात् आप सक्षम होंगे:

- सरल और उन्नत खोज तकनीकों के उद्देश्य का वर्णन करने;
- एक निर्धारित विषय/शीर्ष पर बूलियन सिद्धान्त/तर्क (बूलियन लॉजिक) का प्रयोग करते हुए, सर्च स्ट्रिंग का निर्माण करने;
- एक चित्र की सहायता से सर्च-स्ट्रिंग की व्याख्या करने;



टिप्पणी

- इंटरनेट पर सरल खोज और उन्नत खोज के उदाहरण दे सकने;
- विविध सर्च-इंजनों की पहचान कर सकने, जैसे गूगल, याहू, गूगल स्कॉलर;
- इंटरनेट पर विभिन्न देशी भाषाओं में सर्च इंजनों की पहचान करने;
- विशिष्ट वर्गों में खोज की व्याख्या/करने जैसे मानचित्र, प्रतिरूप (इमेज); और
- सटीक परिणाम प्राप्त करने के लिए सर्च स्ट्रिंग में संशोधन करने में।

18.3 खोज का उद्देश्य

वेब पर सूचना की खोज, लोग विविध उद्देश्यों के लिए करते हैं। सूचना प्राप्त करने के लिए खोज शुरू करने की सर्वश्रेष्ठ विधि है प्रयोगकर्ता की सूचना आवश्यकताओं को परिभाषित करना। सूचना की आवश्यकता, सूचना को ढूँढने और उसे प्राप्त करने की, एक व्यक्ति की अथवा सामूहिक इच्छा है ताकि वह जानी अथवा अनजानी अपनी सूचना आवश्यकता को पूरा कर सके। संभव है किसी शीर्षक (टॉपिक) पर उपयोक्ता की आवश्यकता ही है, उस पर केवल विहंगम दृष्टि डालना, सर्वांगीण खोज, तुरंत संदर्भ अथवा तथ्य, अथवा उस संबंध में गहन-विस्तृत जानकारी प्राप्त करना। एक बार यदि यह निर्णय हो जाये कि आवश्यकता क्या है, तो उस स्रोत का चुनाव किया जा सकता है जिसमें संभवतया वांछित सूचना उपलब्ध हो। तत्पश्चात् खोज रणनीति बनाई जाती है जिसमें विविध स्रोत शामिल हो सकते हैं। वांछित सूचना को खोजने के लिए स्रोतों की एक विशाल श्रृंखला उपलब्ध है।

18.3.1 सर्च इंजन

सर्च इंजन एक सॉफ्टवेयर प्रोग्राम होता है जो खोज-पदावलियों के रूप में संदर्भित पदों के आधार पर वेबसाइटों के लिए खोज क्रियान्वित करता है। इस प्रकार इंटरनेट सर्च-इंजन वेबसाइटों पर उपलब्ध विशिष्ट साइट्स होती हैं जिन्हें अन्य साइट्स में भंडारित सूचना ढूँढने के लिए सहायता देने के लिए निर्मित किया गया है। विविध सर्च इंजन किस प्रकार कार्य करते हैं उनमें अंतर होते हैं, परन्तु वे सभी तीन बुनियादी कार्य करते हैं:

- महत्वपूर्ण शब्दों के आधार पर इंटरनेट पर खोज अथवा इंटरनेट पर उसके किन्हीं भागों (पार्ट्स) का चयन करना,
- उन शब्दों का अनुक्रमणी (इंडेक्स) बनाना, जिन्हें वे ढूँढते हैं और जहाँ वे मिलते हैं, और
- उस अनुक्रमणी में ढूँढे गये शब्दों अथवा शब्दों के संयोग को उपयोक्ताओं द्वारा उपयोग करने की अनुमति देना।

18.3.2 वेब खोज के प्रकार

सूचना आवश्यकता को हम इस प्रकार परिभाषित कर सकते हैं, “किसी व्यक्ति अथवा समूह की कोई सूचना ढूँढने और उसे प्राप्त करने की इच्छा ताकि वे जानबूझकर अथवा अनजाने में

अपनी आवश्यकता की संतुष्टि कर सकें।” सूचना प्राप्ति के लिए यह सबसे पहली आवश्यकता ही सबसे पहली सूचना पुनः प्राप्ति प्रणाली का प्रयोग करने के लिए उसे प्रेरित करती है। वेब खोज के लिए मानी गई प्रतीति आवश्यकता तीन प्रकार की हो सकती है:

- (1) नेविगेशनल खोज
- (2) सूचनात्मक खोज
- (3) परस्पर लेन-देन संबंधी खोज

(1) नेविगेशनल सर्च

नेविगेशनल खोज, वह खोज है जिसमें खोजकर्ता किसी विशिष्ट वेबसाइट पर अथवा किसी विशिष्ट वेबसाइट के किसी वेब पृष्ठ पर जाना चाहता है। अन्य शब्दों में कहें तो यहां खोजकर्ता किसी वेबसाइट पर जाने के लिए वेब सर्च इंजन का प्रयोग करता है। उदाहरण के लिए यदि आप "President of India" की वेब साइट पर जाना चाहते हैं। ऐसा करने के लिए किसी सर्च-इंजन (जैसे गूगल) पर जाकर सिर्फ President of India टाइप करें और वेब में खोज करें। गूगल द्वारा उपलब्ध कराई गई सूची में अन्य लिंक्स के साथ-साथ भारत के राष्ट्रपति के लिंक शामिल हैं। मात्र उस लिंक पर क्लिक करने से ही आप उस वेबसाइट पर पहुँच जाएंगे। इस खोज का परिणाम चित्र 18.1 में दिया गया है।



चित्र 18.1 'प्रेजिडेंट ऑफ इंडिया' शीर्षक पर गूगल सर्च में की गई खोज का परिणाम





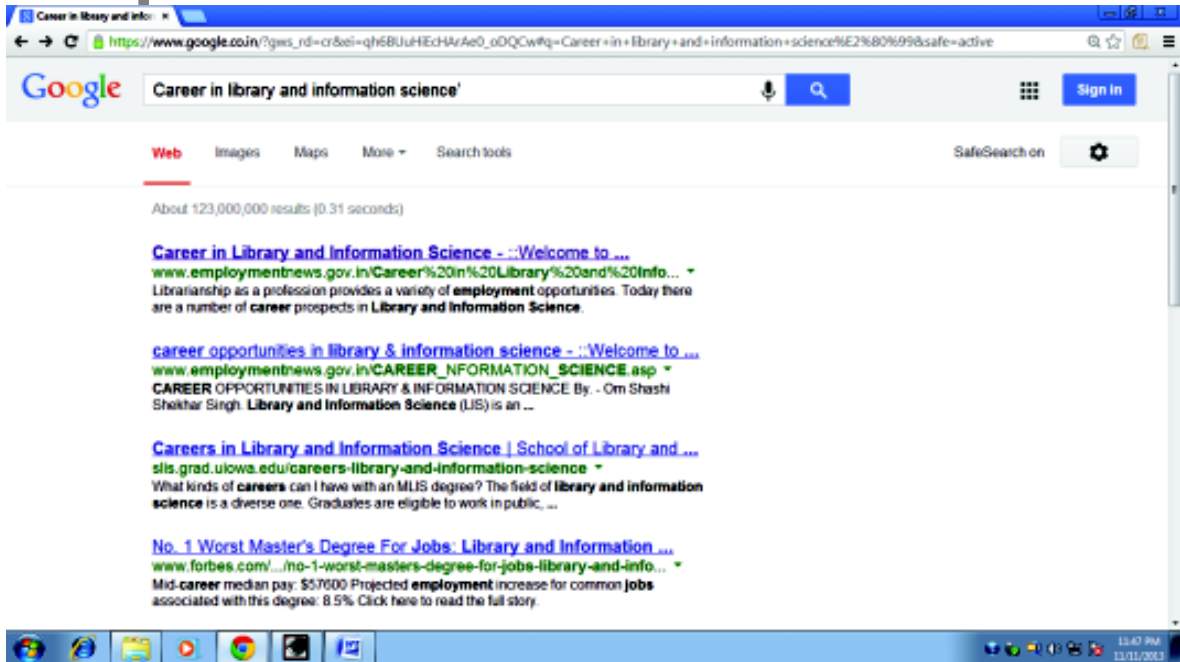
टिप्पणी

(2) सूचनात्मक खोज

सूचनात्मक खोज से आशय है कोई सूचना प्राप्त करना, इस संभावना को मानते हुये कि यह सूचना इंटरनेट पर उपलब्ध है। इस प्रकार की खोज अध्ययन, शोध कार्य अथवा किसी अन्य उद्देश्य के लिए जहाँ विद्वतीय सूचना अपेक्षित हो। उदाहरण के लिए कोई व्यक्ति "Career in Library and Information Science" शीर्षक से संबंधित सूचना ढूँढना चाहता है। जब यह प्रश्न वेब पर गूगल सर्च इंजन पर डालते हैं, तो खोज परिणाम-स्वरूप जो सूची प्राप्त होती है उसमें संपूर्ण वेब में से प्राप्त 42,400 संदर्भ शामिल हैं। यह परिणाम, चित्र 18.2 में दर्शाया गया।

परिणामों के विश्लेषण के बाद, हमें यह पता लगता है कि इस सूची में विभिन्न वेबसाइटों के संदर्भ शामिल हैं। यह वेब साइटें शैक्षणिक, शोध, व्यावसायिक और अनेक अन्य संस्थाओं सहित अनेक संगठनों से संबद्ध हैं।

विशिष्ट सूचना प्रणालियों की खोज, जैसे LCOC, PubMed, विशेष ग्रंथालय का OPAC अथवा इसी प्रकार की खोज प्रणालियों इसी वर्ग के अंतर्गत आती है।



चित्र 18.2 : "Career in Library and Information Science" विषय पर गूगल खोज पर खोज परिणाम

(3) परस्पर लेन-देन संबंधी खोज

परस्पर लेन-देन संबंधी खोज का आशय होता है आगे और परस्पर व्यवहार करने अथवा किसी अन्य गतिविधि को अंजाम देने के लिए किसी वेबसाइट पर पहुँचना इस प्रकार के प्रश्नों का उद्देश्य यह हो सकता है जैसे शॉपिंग विविध प्रकार की फाइलों को डाउनलोड करना, चित्रों,



गानों, चलचित्रों इत्यादि के रूप में और वेब के माध्यम से उपलब्ध करवाई जाने वाली सेवाएँ जैसे गेमिंग इत्यादि। उदाहरण के लिए कोई व्यक्ति उन वेबसाइटों पर जा सकता है जहाँ से वह वायुयान, रेलगाड़ी, बस, चलचित्र इत्यादि के लिए ऑनलाइन टिकटें खरीद सकता है।



पाठगत प्रश्न 18.1

1. सर्च इंजन की परिभाषा लिखें।
2. इंटरनेट पर सूचनात्मक खोज के उद्देश्य का वर्णन करें।

18.4 इंटरनेट पर सूचना खोजना

इंटरनेट पर सूचना की खोज का कार्य कला और विज्ञान दोनों ही है। संबंधित सूचना इंटरनेट पर ढूँढने के कार्य में कोई भी व्यक्ति दक्षता प्राप्त कर सकता है। यह सूचना साधारण वेबसाइटों, डाटाबेसों, पुस्तकों, मानचित्रों, पत्रिकाओं के लेखों, श्रव्य-दृश्य सामग्रियाँ मल्टीमीडिया अथवा वांछित शीर्षकों के रूप में ढूँढी जा सकती है। क्योंकि इंटरनेट पर उपलब्ध सूचना अनेक प्रारूपों में उपलब्ध होती है, इसलिए किसी विशिष्ट खोज तकनीक के माध्यम से अथवा किसी एक सर्च इंजन की सहायता से खोज संभव नहीं है। पाठ 17 में अपने खोज प्रक्रिया, खोज तकनीकों के विभिन्न पक्षों, सर्च स्ट्रिंग अथवा प्रश्नावली इत्यादि की संरचना के संबंध में पढ़ा। इस पाठ में हम, इंटरनेट पर सूचना खोजने के लिए इन पक्षों का अनुप्रयोग करेंगे। सूचना की खोज हम निम्नलिखित सरल अथवा उन्नत तकनीकियों की सहायता से कर सकते हैं।

18.4.1 सरल खोज तकनीक

सरल खोज तकनीक, इंटरनेट पर सूचना की खोज करने वाले बिल्कुल नए खोजकर्ताओं के लिए बनाई गई क्रियाविधि है। इसमें हम मुख्य शब्द का चुनाव करते हैं और बिल्कुल सरल स्ट्रिंग अथवा प्रश्नावली की संरचना करते हैं। जब किसी सर्च इंजन में इस की प्रविष्टि की जाती है तो सर्च इंजन ऐसी सभी संदर्भ सूचियाँ उपलब्ध करवाता है जिनके विषयों में यह मुख्य शब्द मौजूद होता है। यह मूल शब्द प्रलेख के किसी भी भाग में अर्थात् शीर्षक में, पाठ्य वस्तु के बीच में अथवा और कहीं भी मौजूद हो सकते हैं। इंटरनेट पर बहुत विशाल मात्रा में सूचना उपलब्ध होने के कारण सरल खोज क्रियाविधि में बहुत बड़ी संख्या में संदर्भ उपलब्ध कराये जाते हैं। उदाहरण के लिए, www.google.com पर 'career selection after senior secondary' के द्वारा खोज की गई और परिणाम में 4,460,000 मद्दे प्राप्त हुई। चित्र 18.3 में इस खोज का परिणाम दर्शाया गया है।

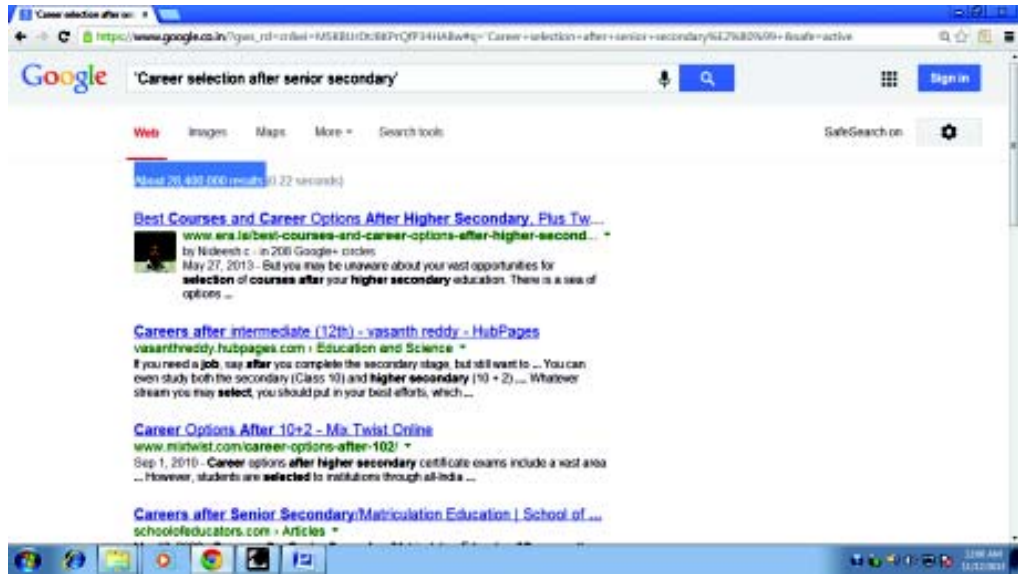
मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



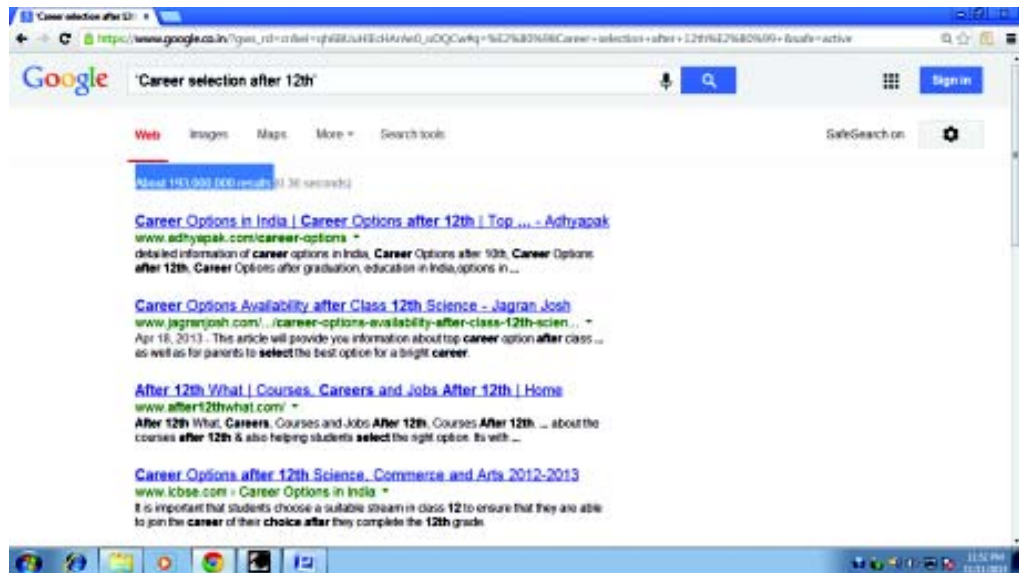
टिप्पणी

सूचना तकनीक : वेब आधारित खोज



चित्र - 18.3 Google पर "Career Selection after Senior Secondary" विषय पर की गई खोज के खोज परिणाम

इसके बाद इसी अवधारणा को "Career Selection after 12th" के रूप में पुनः संरचित किया गया और इसे उसी सर्च इंजन गूगल के माध्यम से खोज क्रियान्वित की गई। इस खोज से लगभग 8,58,000 परिणाम दर्शाये गये। चित्र 18.4 में इस प्रश्नावली के परिणाम दर्शाये गये हैं।



चित्र : 18.4 : "12वीं के बाद जीविका का चयन" विषय पर गूगल सर्च में खोज का परिणाम

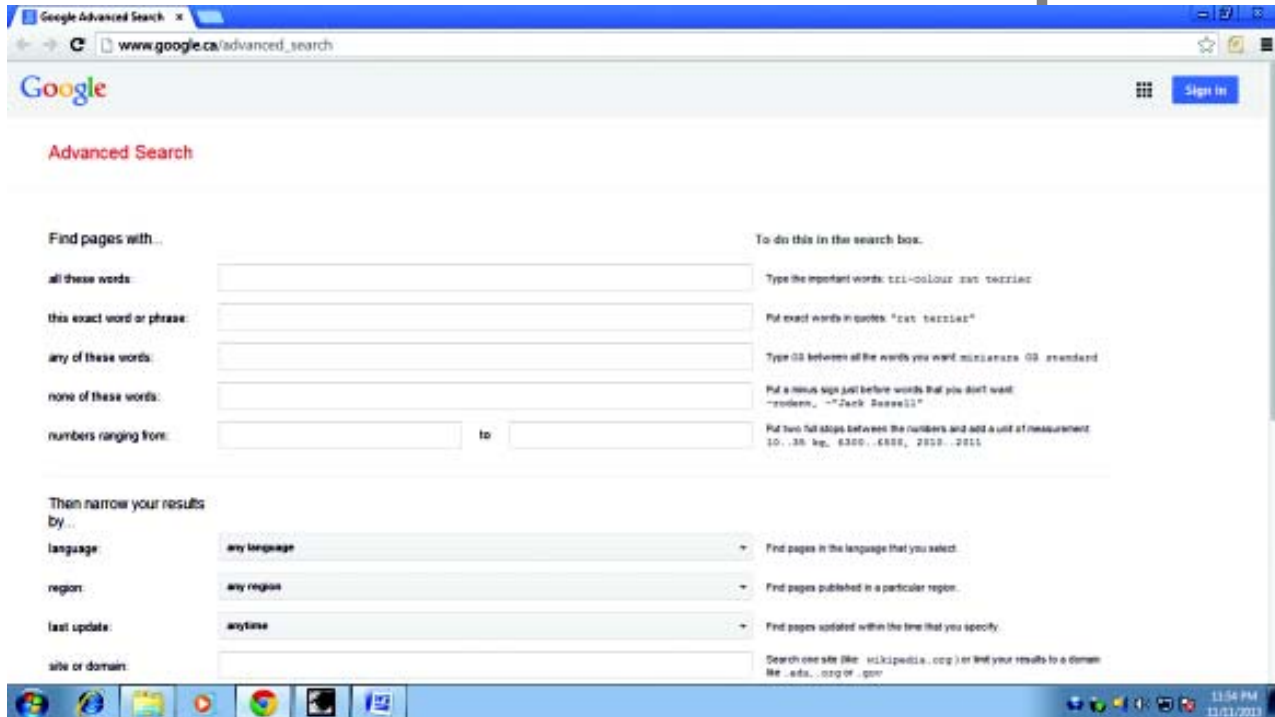


टिप्पणी

इन दो खोजों के परिणामों का विश्लेषण करने के पश्चात् आप यह समझ सकते हैं कि यदि अवधारणायें एक समान भी हों, तब भी यदि मुख्य शब्द में बदलाव करने से खोज के परिणामों में अंतर हो सकता है। अतः खोज के सुस्पष्ट परिणाम प्राप्त करने के लिए खोज की उन्नत तकनीकों के प्रयोग किए जाने की आवश्यकता है।

18.4.2 उन्नत खोज तकनीक

उन्नत खोज तकनीक, वे खोज क्रियाविधियां हैं, जिसमें सुस्पष्ट खोज परिणाम प्राप्त करने के लिए विभिन्न पैमानों का प्रयोग होता है। पिछले पाठ में आपने पढ़ा कि सुस्पष्ट परिणाम प्राप्त करने के लिए स्ट्रिंग तैयार करने में कौन से विभिन्न संचालकों और पैमानों का प्रयोग होता है। यह सिफारिश की जाती है कि किसी भी खोजकर्ता को किसी विशिष्ट सर्च-इंजन द्वारा उपलब्ध कराये गये मार्गनिर्देशों का पालन करना चाहिये और उसी के अनुसार प्रश्नावली बनानी चाहिये। इसके आगे खोज परिणामों के सेट में से संदर्भों की संख्या कम करने के लिए और अधिक पैमाने इसमें जोड़ने चाहिए। उदाहरण के लिए हमने अपनी खोज प्रश्नावली “उच्चतर माध्यमिक के बाद जीविका का चयन” की खोज करने के लिए <http://www.google.ca/advanced-search>, का चुनाव किया। गूगल एडवॉस, और अधिक संबद्ध परिणाम प्राप्त करने के लिए इसी विषय पर खोज को सीमित रखने के लिए, पैरामीटर उपलब्ध करवाता है। इन परिसीमाओं के आधार पर खोज क्रियानिवत करने पर 379,000 संदर्भ प्रदर्शित हुये। इस प्रश्नावली को चित्र 18.5 में और इसके परिणाम को चित्र 18.6 में दर्शाया गया है।

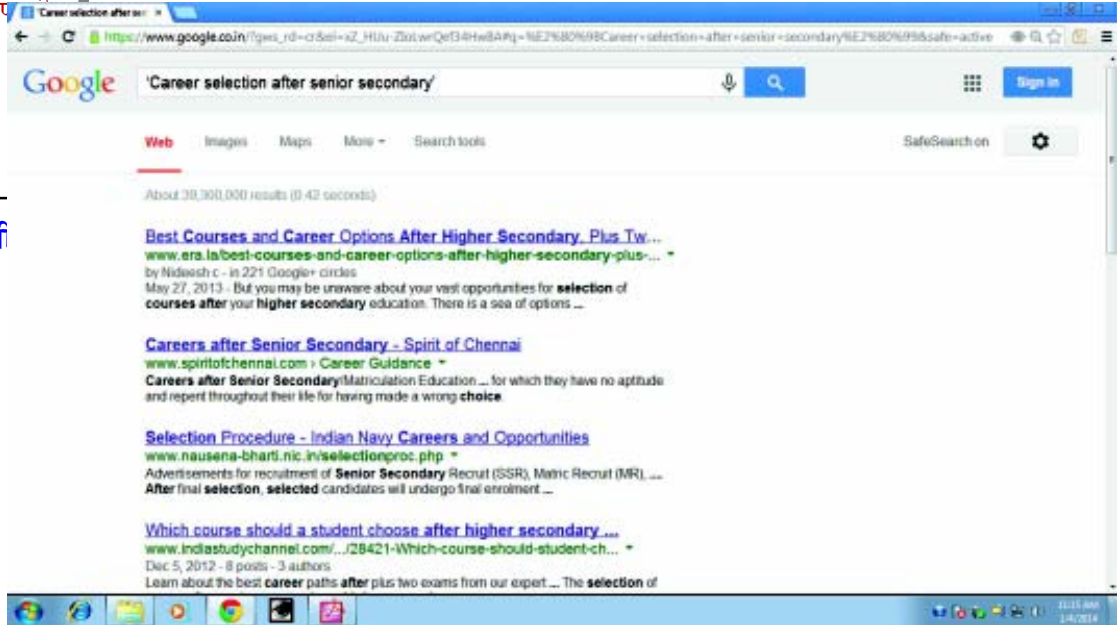


चित्र 18.5 : (“कैरियर सलेक्शन आफ्टर सीनियर सैकेण्डरी” ‘उच्चतर माध्यमिक के बाद जीविका का चयन’ के संबंध में गूगल स्कॉलर पर प्रस्तुत प्रश्नावली

सूचना पुनः प्राप्ति प्रणाली

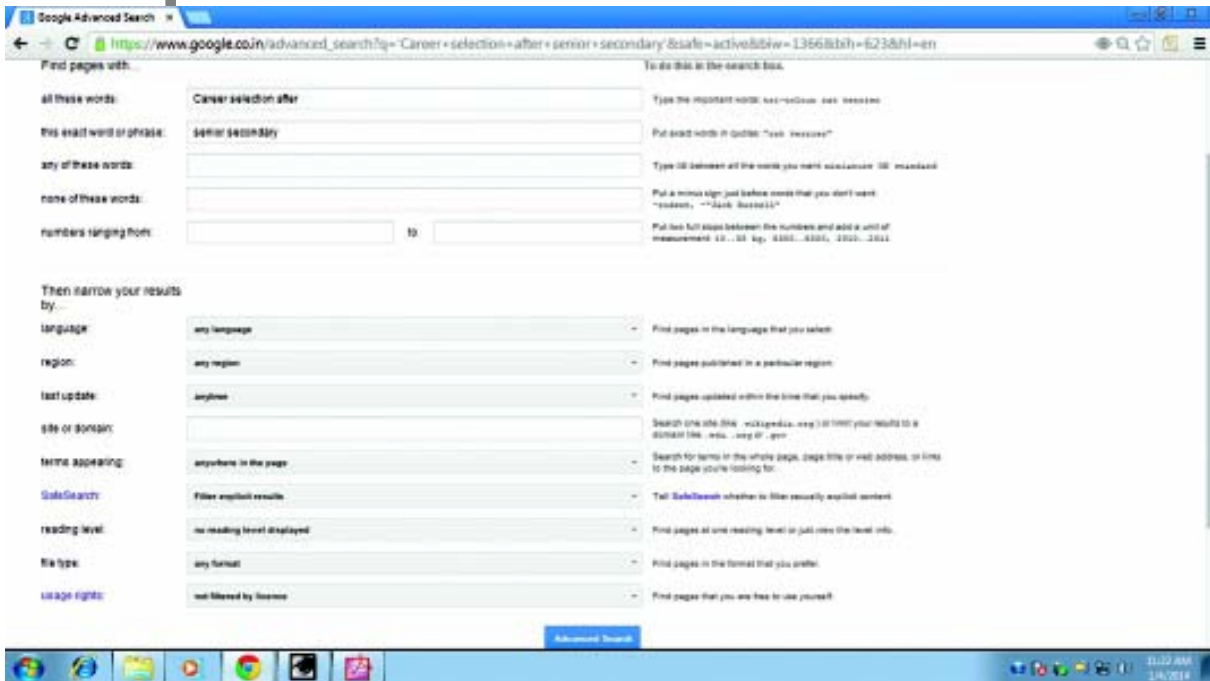


टिप्पणी



चित्र 18.6 : गूगल स्कॉलर द्वारा “कैरियर सिलेक्शन आफ्टर सीनियर सैकेण्डरी” प्रश्नावली के लिए खोज परिणाम

तत्पश्चात् इसी खोज को पुनः तैयार किया गया और, कुछ और पैमाने इसमें जोड़े गये। इस स्ट्रिंग के मुख्य शब्दों वाले इनपुट बॉक्स को चित्र 18.7 में दर्शाया गया है।



चित्र 18.7 : परिसीमाओं का प्रयोग करके गूगल स्कॉलर में प्रस्तुत प्रश्नावली “कैरियर सिलेक्शन आफ्टर सीनियर सैकेण्डरी”



पाठगत प्रश्न 18.2

1. सरल खोज तकनीक की परिभाषा लिखें।
2. उन्नत खोज तकनीक का वर्णन करें।

18.5 बूलियन सिद्धांत और प्रश्नावली

बूलियन लॉजिक और इसके संचालकों के संबंध में पिछले पाठ में पहले ही विवेचना की गई है। आइये अब हम कुछ सर्च स्ट्रिंग्स बनायें और इस लॉजिक का प्रभाव देखने के लिए इन्हें इंटरनेट पर खोजें। बूलियन संचालकों "AND", "OR", "NOT" और <http://support.google.com/websearch/bin/answer.py?hl=en&answer=136861> पर अन्य संचालकों पर प्रयोग के लिए गूगल ने मार्गनिर्देश दिये हैं। गूगल का प्रयोग करने से पहले यह अनुशंसा की जाती है कि बेहतर परिणाम प्राप्त करने के लिए इन मार्ग-निर्देशों को अवश्य पढ़ें।

हमने पुरानी खोज को इस प्रकार बनाया 'career selection after' और 'senior secondary' OR '12th' के साथ भाषा का पैमाना 'english' और क्षेत्र 'India' के अनुरूप तैयार किया। हमें केवल चार संदर्भ प्राप्त हुये। इसका परिणाम चित्र 18.8 में दर्शाया गया है।



चित्र 18.8 : 'career selection after' और 'senior secondary' OR "12th" को 'english' भाषा के पैरामीटर और क्षेत्र "भारत" की प्रश्नावली के खोज परिणाम



टिप्पणी



टिप्पणी



पाठगत प्रश्न 18.3

1. बूलियन लॉजिक में "AND" का क्या प्रयोजन है?
2. बूलियन लॉजिक में "OR" के उद्देश्य को रेखांकित करें?
3. बूलियन लॉजिक में "NOT" के प्रयोजन की व्याख्या करें?

18.6 इंटरनेट पर सर्च इंजन

इंटरनेट पर सैकड़ों सर्च इंजन उपलब्ध हैं। अनेक ऐसी वेबसाइट्स हैं जो सर्च इंजनों के बारे में अध्ययन और विश्लेषण के संबंध में सूचना उपलब्ध करवाती है और वे इंटरनेट पर सक्रिय हैं। आगे सूचना प्राप्त करने के लिए आप <http://www.thesearchenginelist.com/> पर जा सकते हैं। चूंकि सभी 'सर्च इंजनों' और उनकी विशेषताओं को यहाँ उल्लिखित करना व्यावहारिक नहीं है इसलिए कुछे चयनित सर्च इंजनों उनका संक्षिप्त परिचय और उनके वर्गों की एक सूची नीचे दी गई है।

सर्च इंजन

Google™

bing™

YAHOO!

LexisNexis®

dieselpoint

विवरण

Google (गूगल) : विश्व का सर्वाधिक लोकप्रिय सर्च इंजन है।

Bing Search: सर्च इंजन बाजार में माइक्रोसॉफ्ट का प्रवेश देर आपद दुरूस्त आपद।

Yahoo! Search! वेब पर उपलब्ध दूसरा सबसे बड़ा सर्च इंजन (जैसा कि सितंबर 2007 की नीलसन नेटरेटिंग्स रिपोर्ट में परिभाषित किया गया है)

Lexis Nexis : लेक्सिस नेक्सिस का दावा है कि यह "सार्वजनिक अभिलेखों (पब्लिक रिकार्ड्स), अप्रकाशित मत-विचारों, फार्मों, विधिक, समाचारों और व्यावसायिक" सूचनाओं का विश्व का सबसे बड़ा संग्रह है, खोज करने योग्य समाचार-पत्रों के अभिलेख, पब्लिक रिकॉर्ड्स एवं और भी बहुत कुछ इसमें उपलब्ध हैं।

Dieselpoint : खोज और नेविगेशन डीजल-पॉइंट, नेविगेशन सहित, संपूर्ण पाठ्य-वस्तु खोज की क्षमता उपलब्ध करवाता है। यह, उपयोक्ताओं को अत्यधिक



ORACLE
SECURE ENTERPRISE SEARCH **10g**

सटीक के परिणाम देता है जो कि पारंपरिक सर्च इंजन अथवा एस क्यू एल डाटाबेस से भी मिलने संभव नहीं होते।

Oracle Secure Enterprise Search 10g : ऑरेकल से एकमात्र असाधारण उत्पाद, जो संपूर्ण उद्यम सूचना संपदाओं में से सुरक्षित, उच्च गुणवत्ता, प्रयोग में सरल खोज के लिए सक्षमता उपलब्ध कराती है।

SAP

SAP: Net Weaver Search and Classification (TREX) : संरचनात्मक और गैर-संरचनात्मक डाटा में सूचना ढूँढता है। TREX, SAP एप्लीकेशन उपलब्ध कराता है, जिसमें प्रलेखों को खोजने और इन प्रलेखों के वर्गीकरण के लिए सेवाएँ भी होती हैं।

SAIC TeraText
From Science to Solutions

Tera Text Suite : अधिकांश डाटा अर्ध-संरचनात्मक, मूल रूप से पाठ-विषयक (टेक्च्युल) प्रलेखों में उपलब्ध रहता है ना कि संरचनात्मक संगठनात्मक भंडारों (रिपॉजिटरीज) में। Teratext को text-rich डाटा भंडारों के लिए बनाया गया है।

Vivísimo
Clustering Engine

Vivísimo: Clustering Vivísimo clustering Engine : वैज्ञानिकों द्वारा विकसित, बुनियादी इंजन गणितीय कलन पर आधारित जो सर्च टर्म्स में परस्पर संबंधों को जान सके और इसे प्रकाश में लाये (web search : Clusty)

guruji

Guruji.com : India – एक भारतीय इंटरनेट सर्च इंजन जो कि भारतीय संदर्भ में एल्गोरिथम और डाटा के स्वामित्व का लाभ उठाने का भारतीय उपयोक्ताओं को बेहतर खोज परिणाम उपलब्ध करवाने पर संकेन्द्रित है।

rediff.com

Rediff: India – भारत इंटरनेट पोर्टल – समाचारों, संदेशों मेल, मनोरंजन, व्यापार, मोबाइल, खरीदारी, नीलामी, खोज, खेल जगत, आदि।

naukri.com
India's No 1 Job Site

naukri.com (India) : भारत पर आधारित काम (जॉब) ढूँढने का सर्च इंजन।

WebMD
Better information. Better health.

webMD : स्वास्थ्य सूचना लक्षणों की जांच (सिम्पटम चैकलिस्ट) फार्मसी की सूचना का स्रोत और निजी

मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



टिप्पणी

सूचना तकनीक : वेब आधारित खोज

चिकित्सीय सूचना के भंडारण के लिए एक स्थान। यूएस का अंग्रेजी हेल्थ पोर्टल होने के कारण, महीने में इस पोर्टल पर 40 मिलियन खोज करते हैं।



YouTube गूगल की सर्वाधिक बड़ी साईट है। यह सर्च इंजन केवल YouTube के वीडियो के माध्यम से ही खोज करता है।



Find Sounds : किसी भी प्रकार की ध्वनि (साउंड) फाइल ढूँढने के लिए : WAV, MP3, AIFF, AU - सैम्पल रेट और क्वालिटी के द्वारा खोज.... इन ध्वनि प्रभावों को ढूँढने के लिए बहुत ही बढ़िया स्थान।



Ask Jeeves का निर्माण उपयोक्ताओं द्वारा प्राकृतिक भाषा में, दिन-प्रति-दिन की दिनचर्या में पूछे गये प्रश्नों के उत्तर पाने के लिए किया गया था। Ask.com प्रश्न-उत्तर देने वाला वेब का पहला व्यावसायिक सर्च इंजन था।



पाठगत प्रश्न 18.4

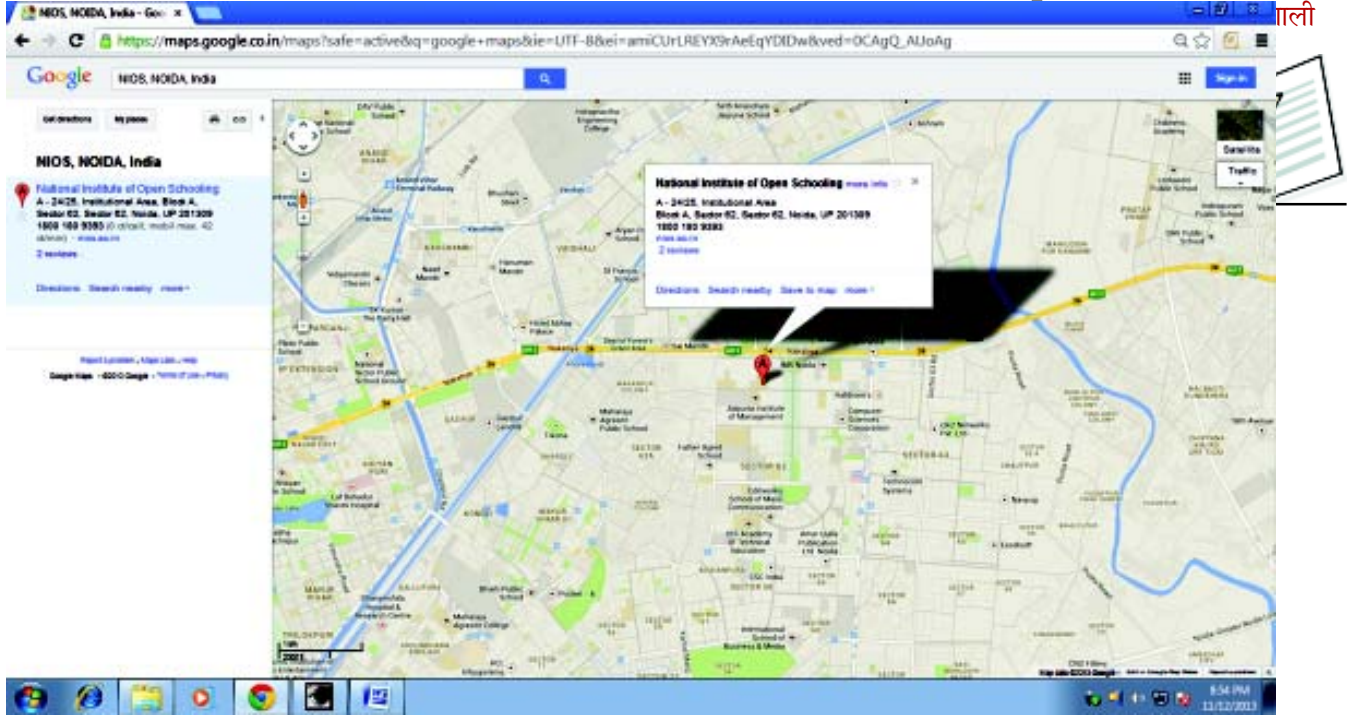
1. WebMD की क्या विशेषतायें हैं?
2. Ask Jeeves सर्च इंजन का वर्णन करें।
3. Dieselpoint की विशेषताएँ लिखें।

18.7 सर्च इंजन और संसाधनों की श्रेणियाँ

इंटरनेट पर सर्च इंजनों ने सूचना स्रोतों को उनके प्रकार और प्रारूप के आधार पर वर्गीकृत किया है। उदाहरण के लिए, गूगल सूचना के स्रोतों को मानचित्रों, छवियों (इमेजिस), समाचारों, विद्वानों, विद्वानों के शोधपत्र, और ऐसे ही अन्य अनेक वर्गों में वर्गीकृत करता है। इन वर्गों का विस्तृत विवरण <http://www.google.co.in/intl/en/about/products/> पर उपलब्ध है। याहू ने सूचनाओं को वेब, वीडियो, न्यूज और स्थानीय वर्गों में वर्गीकृत किया है।

इन वर्गों के अधीन वेब पर सूचनाएँ ढूँढना, वर्ग विशिष्ट का प्रयोग करके खोजना, अत्यन्त प्रभावी बन गया है। इंजन के सर्च आउटपुट में इसकी प्रभावशीलता को देखने के लिए, इस विशेषता का प्रयोग करके गूगल के सर्च इंजन पर एक खोज क्रियान्वित की गई। प्रश्नावली की 'NIOS, NOIDA India'. परिणाम नीचे दर्शाये गये हैं:

(क) मानचित्र वर्ग में हमें NIOS, NOIDA, उत्तर प्रदेश, भारत की अवस्थिति का पता मिला।



चित्र 18.9: मैप वर्ग में गूगल के 'NIOS, NOIDA, India' in map category" का गूगल सर्च आउटपुट

(ख) छवि/चित्र (इमेज) वर्ग में खोजने पर हमें उन स्रोतों का खोज परिणाम मिला जिसमें NIOS के चित्र थे।



चित्र 18.10: "NIOS, NOIDA, India" इमेज वर्ग में गूगल का खोज परिणाम

मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



टिप्पणी

सूचना तकनीक : वेब आधारित खोज

(ग) स्कॉलर वर्ग में खोजने पर 'NIOS, NOIDA, India' के 34 परिणाम (हिट्स) प्राप्त हुये। आलेखों (आर्टिकल्स) का मूल्यांकन करने के उपरांत यह देखने में आया कि जो आलेख प्राप्त हुये उनसे टैक्स्ट में यह स्ट्रिंग उपपस्थित था।



चित्र 18.11: "NIOS, NOIDA, India" की स्कॉलर कैटेगरी में गूगल की खोज का परिणाम

(घ) वेब वर्ग के अधीन 'NIOS, NOIDA, India' की खोज करने पर 293,000 परिणाम (हिट्स) सामने आए। इनमें से कुछ संदर्भों का मूल्यांकन करने पर हमने पाया कि इन वेब पृष्ठों में इस स्ट्रिंग के सभी पद सामने आए।



चित्र 18.12 : वेब वर्ग में "NIOS, NOIDA, India" के गूगल सर्च का परिणाम



इन परिणामों के विश्लेषण से यह देखने में आया कि, खोज की आवश्यकतानुसार प्रभावी परिणाम प्राप्त करने के लिए वर्ग आधारित खोज क्रियान्वित की जानी चाहिये। वर्ग का निर्धारण खोज की आवश्यकता के आधार पर किया जा सकता है।

18.8 स्वदेशी भाषाओं में सर्च इंजन

यदि ग्रंथालय के दृष्टिकोण से देखें तो इंटरनेट विश्व का सबसे बड़ा ग्रंथालय है। इंटरनेट पर अंग्रेजी के अतिरिक्त अन्य भाषाओं में सैकड़ों और हजारों सूचना स्रोत उपलब्ध हैं। यूनिकोड (UNICODE) की खोज होने के बाद से विभिन्न भाषाओं में वेब के साथ सामंजस्यपूर्ण प्रलेखों का सृजन किया गया है और इन्हें वेब पर डाला गया है। इसलिए, अब सर्च इंजन स्वदेशी भाषाओं का प्रयोग करके सूचनाएं खोजने की सुविधा भी प्रदान कर रहे हैं। उदाहरण के लिए "Google.co.in, जो कि भारत में डिफॉल्ट पेज है, अंग्रेजी के अतिरिक्त हिन्दी, बंगाली, तेलुगू, मराठी, तमिल, गुजराती, कन्नड़, मल्यालम और पंजाबी भाषाओं में खोज की सुविधाएं उपलब्ध कराता है। इन सर्च इंजनों की प्रभावशीलता के संबंध में मूल्यांकन संबंधित भाषा के लोगों द्वारा किया जा सकता है। गूगल के ऑफिशियल ब्लॉग में यह दावा किया गया है कि यह चालीस भाषाओं में खोज की सुविधाएं उपलब्ध करवाता है। अतः विभिन्न भाषाओं के लोग, अपनी निजी भाषा में सूचनाएं ढूंढने के लिए सर्च-इंजन का प्रयोग कर सकते हैं।



पाठगत प्रश्न 18.5

1. गूगल स्कॉलर में किस प्रकार की सूचनाओं का समावेश है?
2. कुछ ऐसी भारतीय भाषाओं के नाम लिखें जिनमें गूगल खोज कर सकता है?

18.9 इंटरनेट पर प्रभावी खोज

सर्च इंजनों से प्रभावी परिणाम प्राप्त करने के लिए खोजकर्ता को सर्च इंजनों के व्यवहार एवं विशेषताओं का ज्ञान होना चाहिये। सर्च क्वेरी तैयार करने और सूचना खोजने से पहले, संचालकों, संक्षेपन (ट्रैक्शन) चिन्हों और प्रतीकों अथवा किसी भी अन्य संसाधनों को समझना अत्यंत आवश्यक है। इस उद्देश्य के लिए, यह सुझाव दिया जाता है कि सर्च, इंजनों द्वारा अपनी वेब साइटों पर उपलब्ध कराये गये, मार्गनिर्देशों को खोजकर्ताओं द्वारा अवश्य पढ़ना चाहिए। उदाहरण के लिए गूगल एक समर्पित वेबसाइट प्रस्तुत करता है जहाँ खोजकर्ताओं के लिए मार्गनिर्देश उपलब्ध किये गये हैं (<http://support.google.com/websearch/hl-en>). "Tips for Search" और "Explore Google Search" दो महत्वपूर्ण अनुभाग हैं जो खोजकर्ताओं को इंटरनेट पर प्रभावी रूप में सूचना खोजने में सहायता करते हैं। 'yahoo', Ask, Bing इत्यादि भी खोजकर्ताओं को प्रभावी ढंग से खोज करने में सहायता करते हैं। नीचे दिये गये रेखा चित्र की सहायता से हम इंटरनेट पर खोज प्रक्रिया की भलीभांति व्याख्या कर सकते हैं।

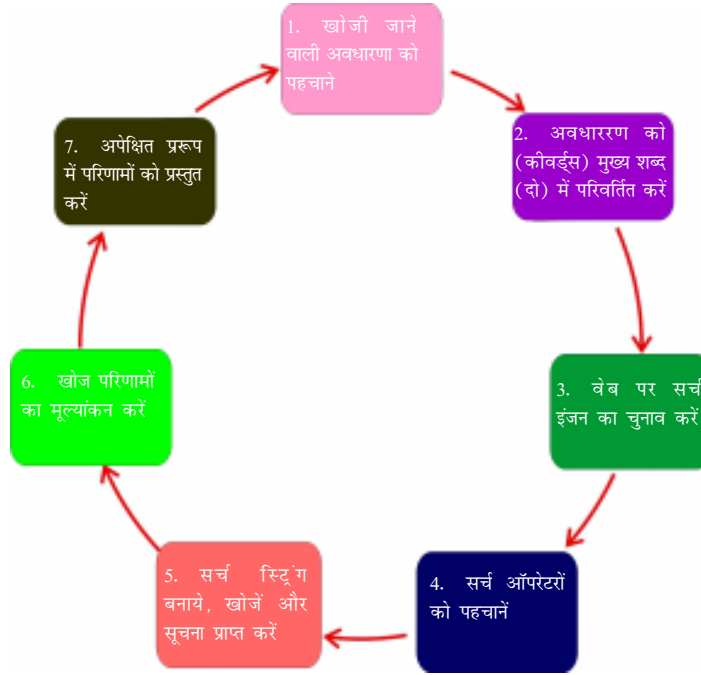
मॉड्यूल-5B

सूचना पुनः प्राप्ति प्रणाली



टिप्पणी

सूचना तकनीक : वेब आधारित खोज



चित्र 18.13 : इंटरनेट पर खोज प्रक्रिया का रेखा चित्र

चित्र 18.13 यह दर्शाता है कि इंटरनेट पर सूचना खोजने के सात चरण हैं। इन चरणों का विस्तृत वर्णन इस प्रकार है:

चरण-1 : उस पूरी अवधारणा की पहचान करें, जिसके संबंध में आप इंटरनेट पर खोज करना चाहते हैं। कुछ संबद्ध मुख्य शब्द, विषय के व्याख्यायों, और/अथवा वाक्यों, और जिन सूचना स्रोतों को आप ढूंढना चाह रहे हैं उनका चुनाव करें।

चरण-2 : अवधारणा, पदों, मुहावरों, उनके विकल्पों और सहयोजित पदों अथवा विषयों इत्यादि को मुख्य शब्द में अनुदित करें।

चरण-3 : खोजे जाने वाले सूचना क्षेत्र (डोमेन) का चुनाव करें, जैसे ग्रंथालय, OCLC, LISA, Pubmed, इत्यादि अथवा इंटरनेट पर सुसंगत सर्च इंजन।

चरण-4 : सूचना क्षेत्र के सर्च इंजन द्वारा प्रयोग में लाये जा रहे संचालकों और प्रतीकों की पहचान करें।

चरण-5 : And/Or संयोजकों के ऑपरेटरों का प्रयोग करके एक सर्च स्ट्रिंग बनायें और खोज क्रियान्वित करें।

चरण-6 : खोज परिणाम का मूल्यांकन करें और यदि अनिवार्य हो तो वैकल्पिक पदों का चुनाव करके खोज में संशोधन करें। विषय अथवा शीर्षक के प्रसार क्षेत्र (स्कोप) को बढ़ायें और सर्च इंजन में उपलब्ध अन्य परिसीमाओं जैसे प्रकाशन की तारीख, भाषा, प्रारूप, स्रोत इत्यादि को भी साथ में जोड़ें।



टिप्पणी

चरण-7 : उपयोक्ता के लिए सहज प्रारूप (फॉर्मेट) में परिणामों को प्रस्तुत करें।

ये सात चरण किसी भी खोजकर्ता को इंटरनेट से सूचना ढूंढने के कार्य में दक्ष/कुशल बना सकते हैं। सर्च क्वेरी के साथ प्रयोग करने, क्वेरी के संयोजन में क्रमपरिवर्तन करने, प्रत्येक क्वेरी के लिए प्रतीकों के समूह का प्रयोग और परिणामों के विश्लेषण के साथ-साथ सर्च स्ट्रिंग में बदलाव करने का प्रभाव किसी भी खोजकर्ता को इंटरनेट पर खोज करने में कार्यकुशल बना सकता है। हमेशा यह सिफारिश की जाती है कि प्रभावी और सुसंगत परिणाम प्राप्त करने के लिए, किसी भी खोजकर्ता को कम से कम दो सर्च इंजनों का प्रयोग करना चाहिए।



आपने क्या सीखा

- इंटरनेट, विश्व में सूचना का सबसे बड़ा भंडार बन चुका है। इसमें हर प्रकार की सूचना डिजिटल रूप में भंडारित की जाती है।
- इंटरनेट पर सूचनाएं खोजने के तीन प्रमुख उद्देश्य हैं, जैसे (क) नैविगेशनल सर्च (ख) सूचनात्मक खोज, और (ग) व्यावसायात्मक (ट्रांजैक्शनल) खोज।
- इंटरनेट पर सूचनाएं खोजना कला और विज्ञान दोनों श्रेणियों में आता है।
- सरल खोज तकनीक इंटरनेट पर नये खोजकर्ताओं द्वारा सूचना खोजे जाने की क्रियाविधि है जिसमें खोजकर्ता किसी सर्च इंजन का चुनाव करता है और उसमें अपना सरल प्रश्न डालता है और उस सूचना की खोजता है।
- उन्नत खोज तकनीक वह क्रियाविधि जिसमें व्यक्ति सुसंगत परिणाम प्राप्त करने के लिए सर्च इंजन द्वारा उपलब्ध कराये गये अन्य ऑपरेटर्स और पैमानों/(परिसीमाओं) (पैरामीटर्स) का प्रयोग कर सकता है।
- बूलियन सिद्धांत (लॉजिक) 'AND', 'OR', 'NOT' को, ऑपरेटर्स की भांति प्रयोग में लाता है, ताकि स्ट्रिंग के पदों में, खोजी जाने वाली अवधारणा को, इस प्रकार व्यक्त किया जाए कि यह सर्च इंजन की समझ में आ सके।
- इंटरनेट पर असंख्य सर्च इंजन उपलब्ध हैं। खोज की अपेक्षानुसार व्यक्ति सर्च इंजन का चुनाव कर सकता है।
- बेहतर परिणाम प्राप्त करने के लिए, सर्च इंजनों ने सूचना स्रोतों को विभिन्न वर्गों में वर्गीकृत किया हुआ है जैसे मानचित्र चित्र/छवियाँ (इमेजिस), पुस्तकें, आलेख (आर्टिकल्स) इत्यादि।
- इंटरनेट पर अंग्रेजी भाषा से इतर अन्य भाषाओं में सूचना स्रोत उपलब्ध हैं। इसलिए, भारतीय भाषाओं में भी सर्च इंजन उपलब्ध हैं, जैसे तमिल, मलयालम, बंगाली इत्यादि।



टिप्पणी



पाठांत प्रश्न

1. इंटरनेट पर सर्च इंजन की क्या आवश्यकता है?
2. इंटरनेट पर खोज के उद्देश्य की विवेचना करें।
3. उन्नत खोज तकनीकियों की विशेषताओं के संबंध में चर्चा/विवेचना करें।
4. बूलियन ऑपरेटर किस प्रकार कार्य करते हैं?
5. स्वदेशी भाषा में सर्च इंजनों से आपका क्या तात्पर्य है?
6. इंटरनेट पर प्रभावी खोज के विभिन्न चरणों का वर्णन करें।



पाठगत प्रश्नों के उत्तर

18.1

1. सर्च इंजन एक सॉफ्टवेयर है, जो उन शब्दों के आधार पर साइटों की खोज करता है, जिनको हम खोज पदों (सर्च टर्म्स) के रूप में संदर्भित करते हैं। इस तरह इंटरनेट सर्च इंजन वेब पर उपलब्ध विशिष्ट स्थान (स्पेशल साइट्स) हैं जिन्हें उन साइटों पर संग्रहित सूचनाएँ खोजने के लिए लोगों की सहायता हेतु बनाया गया है। विभिन्न सर्च इंजनों के कार्य करने के तरीकों में अन्तर होता है।
2. एक तदर्थ (एड हॉक) खोज के लिये खोजकर्ता का लक्ष्य होता है, किसी शीर्षक से संबंधित अधिक से अधिक संभावित सुसंगत प्रलेख ढूँढना/तदर्थ खोज सूचनात्मक प्रकृति की होती है, क्योंकि कोई भी खोजकर्ता किसी विषय के विस्तृत सूचना स्रोतों के संबंध में सूचना ढूँढ़ रहा होता है। उस खोजकर्ता को खोजने वाले विषय के संबंध में पूर्व-जानकारी हो भी सकती है और नहीं भी, परन्तु वह इस विषय पर और अधिक पढ़ना अथवा जानना चाहता है। सूचनात्मक खोज का आशय किसी सूचना को प्राप्त करना है, इस संभावना के साथ कि वह सूचना इंटरनेट पर उपलब्ध होगी।

18.2

1. सरल खोज तकनीक, सरल शब्द के रूप में सर्च इंजन में प्रस्तुत किए गये पदों के आधार पर, इंटरनेट पर सूचनाएँ ढूँढ़ने की क्रियाविधि है।
2. उन्नत खोज तकनीक, खोज की ऐसी क्रियाविधि है जहाँ सुसंगत खोज परिणाम प्राप्त करने के लिए विभिन्न पैमानों (पैरामीटर्स) का प्रयोग किया जाता है।

18.3

1. 'AND' संचालक (ऑपरेटर) का उद्देश्य है उस सूचना स्रोत को ढूँढ़ना, जहाँ इस ऑपरेटर द्वारा जुड़ने वाले दोनों पद मौजूद होंगे।

2. 'OR' ऑपरेटर का उद्देश्य है उस सूचना स्रोत को ढूँढना जहाँ इस ऑपरेटर से संबंधित दोनों पदों में से कोई एक पद मौजूद होगा।
3. 'NOT' ऑपरेटर का उद्देश्य है ऐसा सूचना स्रोत ढूँढना जहाँ पहला पद तो मौजूद होगा परन्तु दूसरा पद नहीं।

18.4

1. WebMD, स्वास्थ्य पाने, लक्षणों की जांच-सूची पाने, फार्मसी सूचना प्राप्ति, और निजी चिकित्सीय सूचनाएँ भंडारित करने के स्थान से संबंधित स्रोत है। यूएस के अग्रणी हैल्थ पोर्टल होने के नाते, इस पर हर महीने 40 मिलियन हिट होते हैं।
2. Ask Jeeves को दिन प्रति-दिन के जीवन की स्वाभाविक भाषा में पूछे गये प्रश्नों के उत्तर पाने के लिए बनाया गया है। Ask.com वेब के लिए अपनी तरह का पहला व्यावसायिक प्रश्न-उत्तर सर्च इंजन था।
3. Dieelpoint, डाटा नैविगेशन की क्षमता सहित, उन्नत संपूर्ण पाठ्य-वस्तु (फुल-टैक्स्ट) खोज उपलब्ध करवाता है। यह सर्च इंजन अत्यधिक सटीक परिणाम उपयोक्ताओं को उपलब्ध करवाता है जिन्हें पारंपरिक सर्च इंजनों अथवा SQL डाटाबेस से प्राप्त करना संभव नहीं है।

18.5

1. गूगल स्कॉलर में विद्वतापूर्ण साहित्य का संपूर्ण पाठ्य उपलब्ध होता है।
2. जिन भारतीय भाषाओं में गूगल खोज कर सकता है, वे हैं, हिंदी, बंगाली, गुजराती, मराठी, कन्नड़, पंजाबी इत्यादि।

पारिभाषित शब्दावली

ब्राउज (Browse) : वेब पृष्ठ के माध्यम से अवलोकन (ब्राउज) करना कि वहाँ क्या है और वह लिंक आपको कहां ले जाता है। जब आप ब्राउज करते हैं तो आपको यह अनुमान लगाना होता है कि कौन से शब्द और लिंक, उस पृष्ठ पर आपकी अभिरूचि से मेल खाते हैं।

ब्राउजर्स (Browser) : वे सॉफ्टवेयर कार्यक्रम जो आपको इंटरनेट पर उपलब्ध वेब पृष्ठों और अन्य प्रलेखों को अवलोकन करने के योग्य बनाता है। वे HTML-encoded फाइलों को पाठ्य-वस्तु, चित्रों, ध्वनियों और अन्य ऐसी विशेषताओं के रूप में अनूदित करते हैं जिन्हें आप देख सकते हैं। सामान्य रूप से प्रयोग किए जाने वाले सर्वाधिक लोकप्रिय ब्राउजर हैं माइक्रोसॉफ्ट (Microsoft), इंटरनेट एक्सप्लोरर, फायरफॉक्स (Firefox), मोजिल्ला (Mozilla), सफारी (Safari) और क्रोम (Chrome)।

हाइपर टैक्स्ट मार्कअप लैंग्वेज (HTML : Hyper Text Markup Language)

इंटरनेट सर्विस प्रोवाइजर (ISP) : वह कंपनी जो इंटरनेट तक सीधे अभिगम (direct access) को बेचती है।

ग्रंथालय एवं सूचना विज्ञान





टिप्पणी

लाइब्रेरी ऑफ कांग्रेस ऑनलाइन कैटलॉग (LCOC):

मेटा-सर्च इंजन (Meta-Search Engine) : वे सर्च इंजन जो अपने आप ही अनेक अन्य सर्च संसाधनों पर आपके मुख्य शब्द को प्रस्तुत कर देते हैं, और उनके संपूर्ण डाटाबेस में से परिणामों की पुनः प्राप्ति कर देते हैं।

ओपैक (OPAC) : ऑनलाइन पब्लिक एक्सेस कैटलॉग

पबमेड (PubMed) : यह निशुल्क अभिगम के लिए डाटाबेस है मूलतः MEDLINE डाटाबेस जो कि जीवन विज्ञानों (लाइफ साइंसिज) और जैव-चिकित्सीय शीर्षों (बायोमैडिकल टॉपिक्स) के संदर्भों और संक्षिप्त प्रलेखों (एबस्ट्रैक्ट्स) का डाटाबेस है।

सरवर/वेब सरवर (Server/Web Server) : कम्प्यूटर पर संचालित होने वाला वह सॉफ्टवेयर, जिसे IP पता दिया गया है और जो इंटरनेट से जुड़ा हुआ है ताकि यह वर्ड वाइड वेब (world wide web) के माध्यम से प्रलेख उपलब्ध करवा सके।

साइट/वेब-साइट (Site/web-site) : इस पद (टर्म) को “वेब पेज” के अर्थ में प्रयोग किया जाता है। वेब पृष्ठ एक एकल क्षेत्र (एंटीटी), एक URL, एक फाइल है जिसे आप संभवतः वेब पर ढूँढ सकते हैं। और अधिक सुस्पष्ट रूप में कहें तो "Site" वह अवस्थान (लोकेशन) अथवा वह समूह (गैदरिंग) या केन्द्र है, उन संबद्ध पृष्ठों के बंच का जो उस साइट की संरचना के लिए परस्पर जुड़े हुये (लिंकड) हैं।

यूनिकोड (UNICODE) : स्टैण्डर्ड कोडिंग स्कीमों का एक सैट, जिसका आशय है संपूर्ण विश्व में वर्तमान में प्रयुक्त की जा रही अनेक (मल्टिपल) कोडिंग स्कीमों को प्रतिस्थापित करना। यूनिकोड कंसौरटियम ने 1991 में, मौलिक मानक, यूनिकोड ट्रांसफॉर्मेशन फॉर्मेट-16 (UTF-16), का मानक कोडिंग स्कीम के रूप में विकास किया ताकि देवनागरी (हिन्दी), बंगाली, चीनी, जापानी, कोरियन इत्यादि, बहुरूपीय जटिल वर्णावाली भाषाओं को समर्थन दिया जा सके।

यूआरएल (URL) : यूनिफॉर्म रिसोर्स लोकेटर। यह किसी भी वेब प्रलेख का अद्वितीय (यूनीक) पता है।

वर्नाकुलर लैंग्वेज (Vernacular Language) : स्वदेशी भाषा किसी विशिष्ट जनसंख्या, समुदाय अथवा क्षेत्र की स्थानीय (स्वदेशी) भाषा और बोलचाल की भाषा होती है।

www : द वर्ड वाइड वेब (The world wide web)

वेब साइट्स

<http://www.thesearchenginelist.com/>

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

www.swse.org

<http://www.w3.org/standards/semanticweb/>

<http://www.google.com>

<http://www.yahoo.com>