टिप्पणी

15

सूचना पुनः प्राप्ति प्रणाली अवधारणा एवं क्षेत्र

15.1 परिचय

किसी विशिष्ट तथ्य अथवा परिस्थिति से संबद्ध ज्ञान संप्रेषण अथवा ज्ञान प्राप्त होना सूचना है। पुनः प्राप्ति (आईआर) से अभिप्राय है उपस्थित कार्य से संबंधित सूचना को भंडारित सूचना में से ढूंढ कर खोज निकालना। इसके मद्देनजर सूचना पुनः प्राप्ति का संबंध है सूचना के पुनः प्रस्तुतीकरण, भंडारण, सूचना के व्यवस्थापन/सूचना-मदों तक अभिगम से संबद्ध कार्य। सूचना मदों के प्रकारों में यहाँ प्रलेख, वेब पृष्ठ, ऑनलाइन, कैटालॉग, सुव्यवस्थित अभिलेख, मल्टीमीडिया मदें (आइटम्स) शामिल हैं। आई आर के प्रमुख उद्देश्य हैं किसी संग्रह में उपलब्ध प्रलेखों की विषय-वस्तु को सूचीबद्ध करना तथा वांछित प्रलेखों की खोज करना। ग्रंथालय, ऐसी सर्वप्रथम संस्थाओं में से एक हैं जिन्होंने सूचना की पुनः प्राप्ति के लिए आई आर प्रणाली को अपनाया।

इस पाठ में आपको सूचना पुनः प्राप्ति के महत्व, इसकी परिभाषाओं तथा इसके उद्देश्यों से परिचित कराया जायेगा। इस पाठ में आप विषय सुगम्यता की दृष्टि से सूचना की अवधारणा, सूचना पुनः प्राप्ति की प्रक्रिया एवं अनुक्रमणिका (इंडेक्सिंग) के संबंध में भी विस्तार से पढ़ेंगे।

15.2 उद्देश्य

इस पाठ के अध्ययन के बाद आप सक्षम होंगे:

- सूचना पुनः प्राप्ति को परिभाषित करने;
- सूचना पुनः प्राप्ति प्रणाली के महत्व एवं अनिवार्यता को समझने;
- विषय तक पहुँचने की सूचना की अवधारणा की व्याख्या करने;
- सूचना पुनः प्राप्ति की प्रक्रिया की सोदाहरण व्याख्या करने; और

टिप्पणी

- प्राकृतिक, मुक्त एवं नियंत्रित अनुक्रमणीकरण (इंडेक्सिंग) की भाषाओं में अन्तर स्पष्ट करने में।

15.3 सूचना पुनः प्राप्ति (आईआर)

''सूचना पुनः प्राप्ति पदावली का आविष्कार कैल्विन मूर द्वारा 1950 में किया गया था। 1961 में जब से सूचना के व्यवस्थापन के लिए कम्प्यूटरों की शुरूआत हुई है तब से अनुसंधान [(शोध)] समुदाय में इस पद को अत्यन्त लोकप्रियता मिली है। इसके बाद 'सूचना' पुनः प्राप्ति पद का संग्रहित प्रलेख डाटाबेस से संदर्भ ग्रंथपरक सूचना की पुनः प्राप्ति के अर्थ में प्रयोग होने लगा था, परन्तु वे सूचना पुनः प्राप्ति प्रणालियाँ, प्रलेख पुनः प्राप्ति प्रणालियां थीं। इनकी रचना उपयोक्ता की प्रश्नावली से संबद्ध संदर्भ ग्रंथपरक प्रलेखों के विद्यमान होने अथवा विद्यमान न होने के बारे में की गई थी अन्य शब्दों में कहें तो प्रारंभ में आई आर एस की रचना खोज के किसी अनुरोध के प्रत्युत्तर में पूरे प्रलेख (पुस्तक, कोई आलेख इत्यादि) की पुनः प्राप्ति के लिए की गई थी।

यद्यपि आज के आई आर एस यही कार्य करते हैं परन्तु वक्त बीतने के साथ-साथ अनेक उन्मत तकनीकियां विकसित हो गई हैं तथा यह आई आर एस की रचना के लिए अनुपयुक्त बनती जा रही हैं। समय व्यतीत होने के साथ सूचना पुनः प्राप्ति के संकेतार्थ में परिवर्तन आया है तथा इसके लिए सूचना व्यवसायियों व शोधार्थियों द्वारा अलग-अलग पदों का प्रयोग किया जा रहा हैं। इनमें से कुछ पद हैं–सूचना संग्रहण एवं पुनः प्राप्ति, सूचना व्यवस्थापन तथा पुनः प्राप्ति, सूचना को संसाधित करना व इसकी पुनः प्राप्ति, पाठ्य-वस्तु की पुनः प्राप्ति, सूचना प्रस्तुतीकरण व पुनः प्राप्ति एवं सूचना तक अभिगमयता।

आइये अब हम सूचना पुनः प्राप्ति के उन साधनों को समझने का प्रयास करते हैं जिनके माध्यम से ग्रंथालयों और कुछ प्रणलियों के द्वारा उनके संग्रह में उपलब्ध प्रलेखों से सूचना खोजने का कार्य किया जाता है। यदि उपयोक्ता की आवश्यकतानुसार जिस समय और जब आवश्यक हो तब सही प्रलेखो की पुनः प्राप्ति ग्रंथालय से संभव न हो पाये तो ग्रंथालय का संग्रह चाहे जितना भी बड़ा हो उसका कोई महत्व नहीं है। यह कार्य करने के लिए किसी न किसी सूचना पुनः प्राप्ति प्रणाली का होना आवश्यक है। जब मांगी गई सूचना और पुनः प्राप्ति प्रणाली में उपलब्ध सूचना में मेल हो जाये तो अपेक्षित प्रलेखों को ढूंढ लिया जाता है। अन्य शब्दों में कहें तो इसका तात्पर्य है कि उपयोक्ताओं द्वारा मांगी गई सूचना के साथ उपलब्ध कराये गये प्रलेख स्वीकार्य स्तर तक मेल खाते हैं। मेल खाती हुई सूचना को सफलतापूर्वक ढूंढना ही सूचना पुनः प्राप्ति का मुख्य उद्देश्य है।

किसी ग्रंथालय का मुख्य कार्य है उपयोक्ताओं को उनकी आवश्यकतानुसार सूचना उपलब्ध करवाना। इस कार्य को पूरा करने के लिए ग्रंथालय में संग्रहित डाटाबेस से सूचना को पुनः प्राप्त करना आवश्यक है। 'सूचना पुनः प्राप्ति' संग्रहित सूचना भंडार से सूचना चयनित करने की प्रक्रिया है। यह प्रक्रिया धीरे-धीरे कंप्यूटरों और दूरसंचार प्रौद्योगिकी पर अधिकाधिक आश्रित होती जा रही है। सूचना पुनः प्राप्ति प्रणाली का डिजाइन अनुप्रयुक्त सूचना प्रौद्योगिकी का एक महत्त्वपूर्ण क्षेत्र बन गया है।

टिप्पणी

पाठगत प्रश्न 15.1

1. सूचना पुनः प्राप्ति किसी ग्रंथालय का एक महत्त्वपूर्ण कार्य क्यों है?

15.4 सूचना पुनः प्राप्ति प्रणाली

पारिभाषित शब्दावली के दृष्टिकोण से देखें तो सूचना पुनः प्राप्ति प्रणाली (आई आर एस) की अवधारणा स्वतः स्पष्ट है तथा इसका अभिप्राय है- ''वह प्रणाली जो सूचना पुनः प्राप्ति करती है''। आई आर एस का संबंध दो आधारभूत पहलुओं के साथः (1) सूचना का भंडारण कैसे करें, और (2) सूचना की पुनः प्राप्ति कैसे करें।

सरल शब्दों में इसे हम इस प्रकार निरूपित कर सकते हैं–''एक ऐसी प्रणाली जो सूचना का भंडारण एवं उसकी पुनः प्राप्ति करती है। आई आर एस, परस्पर क्रियात्मक घटकों का समुच्चय (सेट) है, जिसके प्रत्येक घटक को किसी विशिष्ट उद्देश्य को पूरा करने वाले किसी विशेष कार्य को पूरा करने हेतु संरचित किया गया है। यह सभी घटक एक लक्ष्य प्राप्त करने के लिए अंतः संबंधित है अतः आई आर की अवधारणा इस तथ्य पर आधारित है कि सूचना की कुछ ऐसी मदे हैं जिन्हें सुगमता से पुनः प्राप्ति के लिए पहले से ही अनुमान लगाकर, उपयुक्त क्रम में व्यवस्थित किया गया है।

किसी भी पुनः प्राप्ति प्रणाली की संरचना सूचना के विश्लेषण, इसके संसाधन और सूचना के भंडारण स्त्रोतों के और किसी विशिष्ट उपयोक्ता की मांग से मेल खाती सूचना के लिए उनके पुनः प्राप्ति के लिए की जाती है। आधुनिक सूचना पुनः प्राप्ति प्रणाली या तो संदर्भ ग्रंथ सूची की मदों को पुनः प्राप्त कर सकती है अथवा उस तथ्य विषय-वस्तु को पुनः प्राप्त कर सकती है जो प्रलेखों के भंडारित डाटाबेस में से उपयोक्ता के खोज के मानकों से मेल खाती हो। मूलरूप से आई आर एस का अभिप्राय था विषय-वस्तु (टेक्स्ट) पुनः प्राप्ति की प्रणालियां, क्योंकि वे सभी विषयपरक प्रलेखों से संबंध रखती थीं। आधुनिक सूचना पुनः प्राप्ति प्रणालियां न केवल मूल विषय-वस्तु से संबंधित है बल्कि सूचना के साथ-साथ इनका संबंध विषय संबंधी सूचना के बाद मल्टीमीडिया सूचना से भी होता है। जिसमें पाठ्य वस्तु, श्रव्य, चित्र और वीडियो भी शामिल होते हैं। अतः आधुनिक सूचना पुनः प्राप्ति प्रणाली का संबंध पाठ्य के भंडारण, व्यवस्थापन और विषय-वस्तु तक पहुँच के साथ-साथ मल्टीमीडिया सूचना स्त्रोतों के साथ भी होता है।

इस प्रकार 'आई आर प्रणाली' निम्नलिखित में से कुछ अथवा सभी संक्रियाओं के निष्पादन के नियमों व प्रक्रियाओं सूचना पुनः प्राप्ति का एक समुच्चय (सेट) है:

क) अनुक्रमणीकरण (प्रलेखों के प्रस्तुतीकरण के लिए प्रस्तुति पद बनाना)

ख) खोज-सूत्र बनाना (सूचना आवश्यकताओं के अनुरूप प्रस्तुतियाँ की रचना करना।

ग) खोज (आवश्यकतानुसार प्रस्तुतियों की मांग के अनुरूप उनसे मेल खाते प्रलेखों की प्रस्तुतियाँ बनाना और

घ) अनुक्रमणीकरण की भाषा बनाना (प्रस्तुतियों के लिए नियमों का निर्माण करना)

अत: समग्र रूप में सूचना पुन: प्राप्ति को हम इस प्रकार परिभाषित कर सकते हैं–''खोज का विज्ञान'' अथवा डेटा की फाइलों में से सूचना लेकर तैयार की गई ऐसी प्रक्रिया, विधि और पद्धति जिसे संग्रहित एवं सूचीबद्ध सूचना के चयन या उसकी पुन: प्राप्ति के लिए उपयोग में लाया जाता है''

15.4.1 आई आर एस के उद्देश्य एवं कार्य

सूचना पुन: प्राप्ति का मुख्य उद्देश्य है, मांगी गई सूचनाएँ आवश्यकतानुसार, पुन: प्राप्त करवाना। यह सूचना वास्तविक हो सकती है अथवा उन प्रलेखों से प्राप्त सूचना हो सकती है, जिसमें स्थानापन्न (सरोगेट) सूचना सम्मिलित हो जो उपयोक्ता की प्रश्नावली से पूरी तरह अथवा आंशिक रूप से मेल खाती हो। अत: इसमें खोज के परिणाम स्वरूप प्रलेखों के ग्रंथात्मक विवरण जो खोज प्रश्न से संबंधित हों, वास्तविक विषय (टेक्स्ट) चित्र, वीडियो इत्यादि शामिल हो सकते हैं जिनमें अपेक्षित सूचना निहित सूचना पुन: प्राप्ति प्रणाली में डाटाबेस की स्थिति में, प्रलेख का संक्षिप्त सार अथवा पूर्ण विषय-वस्तुएँ, जैसे समाचार-पत्र के लेख, हस्त पुस्तिकाएँ (हैंड बुक्स) शब्द कोश, विश्व कोश, विधिक दस्तावेज, सांख्यिकीय आंकड़े इत्यादि के साथ-2 श्रव्य चित्र और वीडियो सूचना भी शामिल हो सकती है।

आई आर एस के प्रमुख कार्य हैं:

1. लक्षित उपयोक्ताओं के समुदाय की रूचि के क्षेत्र संबद्ध सूचना के स्रोतों की पहचान करना;
2. स्रोतों (प्रलेखों) में निहित विषय वस्तु का विश्लेषण करना;
3. उपयोक्ताओं की प्रश्नावालियों से मिलान के लिए विश्लेषित स्रोतों के निहित विषय को पुन: प्रस्तुत करना
4. खोज के द्वारा ढूंढे गये विवरण का भंडारित डेटाबेस के साथ मिलान कराना;
5. संबद्ध सूचना को पुन: प्राप्त करना; तथा
6. उपयोक्ताओं से प्राप्त प्रति पुष्टि (फीडबैक) के आधार पर प्रणाली में आवश्यक समायोजन करना।

पाठगत प्रश्न 15.2

1. सूचना पुन: प्राप्ति प्रणाली (आई आर एस) के प्रमुख उद्देश्य क्या हैं?

15.5 सूचना पुन: प्राप्ति का महत्व

ग्रंथालयों में अनेक प्रकार के भौतिक प्रारूपों में सूचना उपलब्ध होती है; जबकि कई

टिप्पणी

उपयोक्ताओं के लिए पुस्तकें अभी भी सूचना के संप्रेषण के लिए मुख्यवाहक हैं और अन्यों के लिए आवधिक पत्रिकाएँ (पीरियॉडिकल) अथवा तकनीकी रिपोर्ताज (रिपोर्ट्) ने पुस्तक को प्रतिस्थापित कर दिया है और कुछ अन्यों के लिए, फिल्में या ग्रामोफोन रिकार्ड महत्त्वपूर्ण हैं। अतः यह स्पष्ट है कि सूचना स्रोत अनेक भौतिक प्रारूपों में उपलब्ध हो सकते हैं। बौद्धिक विषय-वस्तु प्रत्येक स्थिति में एक समान ही होगी, परन्तु यह स्वतः स्पष्ट है कि विभिन्न भौतिक प्रारूपों को एक साथ व्यवस्थित करने का प्रयास व्यवहारिक नहीं है। इसलिए, समान सूचना स्रोत के विविध प्रारूपों को ढूँढने के लिए हम ग्रंथालय में मदों की भौतिक व्यवस्था पर भरोसा नहीं कर सकते। हमें ग्रंथालय के किसी अन्य विकल्प विषय-वस्तु के अभिलेखों के एक समुच्चय पर भरोसा करना होगा। यह अभिलेख ग्रंथालय की सूचियों और संदर्भ ग्रंथ-सूचियों के रूप में होते हैं। तथापि, ग्रंथालय की प्रसूची ग्रंथालय के प्रलेखों को ढूंढने की सारी युक्तियों में से एकमात्र उपलब्ध कारगर कुंजी है। कोई भी ग्रंथालय अपने संग्रह में बड़ी संख्या में उपलब्ध पत्रिकाओं में प्रकाशित लेखों की सूची बनाने का प्रयास नहीं करेगा। इसके बजाय हम उन अनुक्रमणिकाओं, सारांशों और इसी प्रकार के ग्रंथात्मक साधनों पर निर्भर करेंगे, जैसे कि इसके माध्यम से हम विभिन्न दृष्टिकोणों के जरिये किसी विशिष्ट मद तक पहुँच बनाने में सक्षम होते हैं।

पाठगत प्रश्न 15.3

1. ग्रंथालय में प्रसूचियों तथा ग्रंथात्मक उपकरणों के महत्त्व का वर्णन करें।

15.6 सूचना तक विषयपरक अभिगम

उपयोक्ता प्रायः कोई प्रश्नावली लेकर सूचना स्रोतों तक पहुँचते हैं जिसका उन्हें उत्तर अथवा सूचना अपेक्षित होती है अथवा वे किन्ही विशिष्ट विषयों से संबद्ध प्रलेख से संबंधित सूचना प्राप्त करना चाहते हैं। उपयोक्ताओं द्वारा विभिन्न स्रोतों से इस प्रकार सूचना प्राप्त करने की पद्धति को हम सूचना तक विषयपरक अभिगम कहते हैं। इस प्रकार की सूचना प्राप्त करने के लिए, सूचना संगठनों के लिए अनिवार्य है कि वे प्रलेखों, ग्रंथालय में उपलब्ध वैकल्पिक प्रलेखों अनुक्रमणियों अथवा डाटाबेस को इस प्रकार व्यवस्थित करें कि उन विशिष्ट मदों से संबद्ध सूचना को पुनः प्राप्त किया जा सके। विषय परक अभिगम का उपयोग करके प्रलेखों में समाविष्ट सूचना उपलब्ध करवाने की अनेक पद्धतियाँ हैं। उनमें से दो प्रमुख पद्धतियाँ हैं:

- वर्णाक्रमानुसार विषय परक अभिगम
- विषय का प्रदर्शन संबद्धता

15.6.1 वर्णानुक्रमानुसार विषय परक अभिगम

इसमें सूचना की मदों को पहले विषयों के शीर्षक के अधीन समूहबद्ध किया जाता है और तत्पश्चात् इन विषय शीर्षकों वर्णानुक्रम में इस प्रकार व्यवस्थित किया जाता है कि उन विशिष्ट

विषयों को हम सरलता से पुन: प्राप्त कर सकें। इस प्रक्रिया में जिन समस्याओं को दूर करने की जरूरत है वे समानाकार शब्दों पर्यायवाची, एक वचनीय अथवा बहुवचनीय रूप, जटिल और मिश्रित शब्दों अथवा विषयों और बहुशब्दीय पद की अवधारणा से संबंधित हैं।

15.6.2 विषय संबद्धता का प्रदर्शन

मनुष्यों की तरह, विषयों के भी परस्पर संबंध होते हैं जिनमें वाक्य-विन्यास संबंध और अर्थगत संबंध शामिल हैं। वाक्य-विन्यास संबंध के अंतर्गत शब्दों और वाक्यांशों को व्यवस्थित करने में, आपसी संबंधों को, दर्शाते हैं। उदाहरण स्वरूप मुख्य शब्द 'फोटोग्राफ्स और अलबम्स' के लिए उपयोक्ता को 'फोटोग्राफ्स ऑफ अलबम्स' और 'अलबम्स ऑफ फोटोग्राफ्स' दोनों के लिए सुविधा दी जानी चाहिए। अर्थगत संबंध में वंशानुगत प्रजातियों (जीनस स्पीशिज) के संबंधों को दर्शाते हैं। उदाहरण स्वरूप, मर्करी (ग्रह) और मर्करी (धातु) में वंशानुगत अर्थपरक अन्तर है, यद्यपि दोनों शब्दों का उच्चारण और वर्तनी एक समान है।

विषयों के अनुसार विस्तृत व्यवस्थापन पर विचारण करने वाले सर्वप्रथम ग्रंथालयी मेल्विल ड्यूई। ड्यूई के पूर्व ग्रंथालयी अपने ग्रंथालयों में वर्गीकृत क्रम में व्यवस्था करते थे, वर्गीकृत प्रसूची से सभी परिचित थे। तथापि, वह वर्गीकृत व्यवस्था विषय नामपद का व्यापक रूप था; इसमें ड्यूई द्वारा सुझाये अनुसार विषय के विशिष्ट विनिर्देशन का विस्तृत विवरण देने का प्रयास शामिल नहीं था। लेकिन वह अनिवार्य एवं उपयोगी था। ड्यूई की वर्गीकरण प्रणाली से दो प्रयोजन पूरे होते थे पहला था पटल पर रखी गई पुस्तकों का व्यवस्थापन; और दूसरा, प्रसूचियों और वांडम्मय-सूचियों में की गई प्रविष्टियों का व्यवस्थापन।

विषयपरक अभिगम अथवा विषयपरक अनुक्रमणिकरण पदों शब्दों, शब्द समूहों, वाक्यों वर्गीकरण वर्गों, संकेत लिपि को पहचानने और उनके चयन की एक प्रक्रिया अथवा तकनीक है कि जिसमें यह संकेत मिलता है कि संबद्ध प्रलेख किस बारे में है। इससे उपयोक्ताओं का प्रलेख में समाविष्ट विषय-वस्तु का सारांश बनाने और इसकी पुन: प्राप्ति के अवसरों की वृद्धि में सहायता मिलती है। अन्य शब्दों में कहें तो इसका संबंध है प्रलेखों के विषयों को पहचानने एवं उनका विस्तृत विवरण देने में। इसका उद्देश्य है प्रलेख की विषय वस्तु के आधार पर विशिष्ट सूचना को ढूंढने की सुविधा उपलब्ध करवाना।

विषय अनुक्रमणीकरण के दो चरण है:

क) प्रलेख का निरूपण करने वाली अवधारणाओं को उत्पन्न करने के लिए विषय विश्लेषण; और

ख) पुन: प्राप्ति के लिए अवधारणाओं को नियंत्रित और सटीक शब्दावली में अनूवादित करना।

पाठगत प्रश्न 15.4

1. सूचना तक ''विषयपरक अभिगम'' की अवधारणा कैसे अस्तित्व में आई? व्याख्या करें।

टिप्पणी

15.7 अनुक्रमणीकरण की भाषायें

जैसे कि ऊपर चर्चा की गई है जब विषय तक पहुँचने के लिए सूचना को प्रयुक्त करते है तो उन्हें विषय के अनुक्रमणीकरण (इंडेक्सिंग) के कठिन कार्य से जूझना पड़ता है। उन्हें प्रलेखों की प्राकृतिक भाषा की जटिलता, विविधता और समृद्धता पर विचार करना पड़ता है। अनुक्रमणी बनाने में असीमित अथवा अनियंत्रित शब्द समुच्यों अथवा मुहावरों का प्रयोग, इन प्रयासों को व्यर्थ कर देता है। उपयोक्ताओं द्वारा विस्तृत मात्रा में प्रयोग किये जाने वाले शब्दों के प्रयोग के कारण खोज में असफलता की मात्रा में वृद्धि होती है। सही कहा गया है कि किसी भाषा के दो शब्दों का कभी एक समान अर्थ नहीं होता और कभी भी उनके वास्तविक पर्याय नहीं होते। परन्तु अर्थों के संदर्भ में शब्दों में प्रायः परस्पर सानिध्य होता है और अधिकतर मामलों में इनके अर्थ-भेद को प्रायः स्पष्टतया समझा नहीं जाता। स्थिर/परिवर्त्ती शब्दों का प्रयोग भी असफलता की ओर ले जा सकता है क्योंकि उपयोक्ता संभवतया अनुक्रमणीकार अथवा प्रलेखो के लेखकों द्वारा प्रयुक्त शब्दों का प्रयोग नहीं कर रहा हो। अनुक्रमणीकरण की विविध जटिल समस्याओं को हल करने के लिए अनेक प्रकार की नियंत्रित शब्दावलियां विकसित की गई हैं। प्रलेखों के वृहद् भंडार में से विषय सूचना की पुनः प्राप्ति के लिए अपेक्षित है कि अवधारणाओं की पहचान और व्यवस्था जानने योग्य रूप में हो। अनुक्रमणीकरण, वह अभियांत्रिकी है जिसके द्वारा प्रलेखों में निहित सूचना को व्यवस्थित किया जा सकता है। परन्तु मुख्य समस्या है अवध ारणाओं को पहचानने और उनकी व्यवस्था करने की। अभिलेखीय सूचना में, लेखक उन प्राकृतिक भाषाओं में संप्रेषण करते हैं जिन्हें भाषायी विशेषताओं द्वारा चित्रित किया जाता है। प्राकृतिक भाषाओं की समस्याओं के समाधान के लिए कृत्रिम भाषा अथवा अनुक्रमणीकरण की भाषा की आवश्यकता अनुभव हुई। इसका अभिप्राय यह हुआ कि अनुक्रमणीकरण की भाषा वह भाषा हैं जिसे प्रलेखों के विषय के वर्गीकरण अथवा अनुक्रमणीकरण के लिए प्रयोग में लाया जाता है। अनुक्रमणीकरण की भाषा को हम इस प्रकार परिभाषित कर सकते है कि यह पदों के समूह का वह समुच्चय है जिसे प्रलेखो के शीर्षो अथवा विशिष्टताओं को प्रस्तुत करने और उन पदों को मिलाने (जोड़ते) अथवा उनके प्रयोग के लिए नियम बनाने के लिए प्रयुक्त किया जाता है। अनुक्रमणिका की भाषा का प्रयोजन है प्रलेखों की अवधारणाओं को एक कृत्रिम भाषा में अभिव्यक्त करना ताकि उपयोक्ता अपेक्षित सूचना प्राप्त करने के लिए उपयोक्ता सक्षम हो सकें। विभिन्न संबंधित अवधारणाओं में परस्पर संबंधों को दर्शाकर अनुक्रमणीकरण की भाषा इसे संभावित बनाती है।

तीन प्रकार की प्रमुख अनुक्रमणीकरण भाषायें हैं.

1. **प्राकृतिक अनुक्रमणीकरण भाषाः-** प्रलेख का वर्णन देने के लिए संदर्भित प्रलेख में से लिया गया कोई भी पद प्रयोग में लाया जा सकता है।

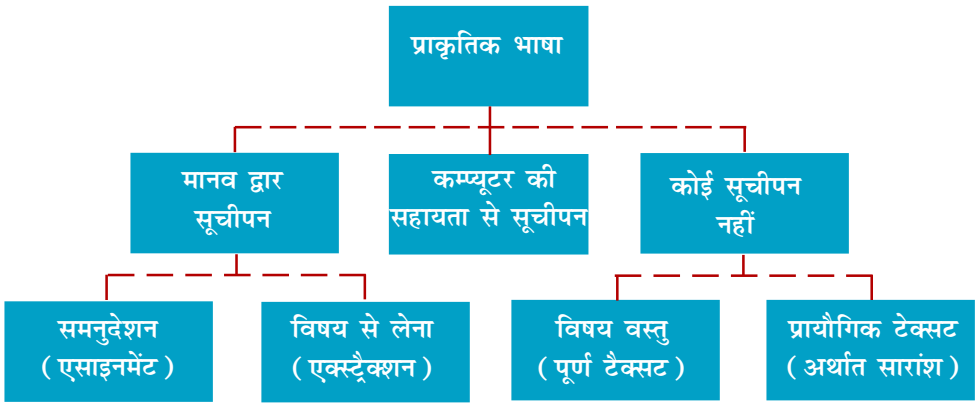
2. **मुक्त अनुक्रमणीकरण भाषाः-** प्रलेख के विवरण के लिए किसी भी पद (ना केवल प्रलेख में से) का प्रयोग किया जा सकता है।

3. **नियंत्रित अनुक्रमणीकरण भाषाः-** किसी प्रलेख का वर्णन करने के लिए सूचीकार द्वारा केवल स्वीकृत पदों का ही प्रयोग किया जाएगा।

निम्नलिखित अनुभागों में संक्षिप्त रूप में आपका परिचय प्राकृतिक, मुक्त और नियंत्रित अनुक्रमणीकरण भाषा से कराया जायेगा।

15.7.1 अनुक्रमणीकरण की प्राकृतिक भाषा

प्राकृतिक भाषा हमारी उस भाषा की ओर इंगित करती है, जिसे हम प्राय: संप्रेषण के लिए प्रयोग में लाते है। जबकि वे भाषायें जिन्हें हम किसी विशिष्ट प्रयोजन के लिए रचते हैं अथवा जिन्हें कोई विशिष्ट अर्थ देने के लिए प्रयुक्त करते हैं अथवा सीमित प्रयोग के लिए प्रयुक्त करते हैं, उसे कृत्रिम भाषा कहते है। अत: अनुक्रमणी की प्राकृतिक भाषाये हैं ''प्राकृतिक भाषा अथवा जिस प्रलेख की अनुक्रमणी तैयार जा रही हो उसकी सामान्य भाषा। कोई भी ऐसा पद जिसका उस प्रलेख में उल्लेख होता है वह अनुक्रमणी के पदों के लिए प्रत्याशी पद होता है। व्यवहार में, प्राकृतिक भाषा में अनुक्रमणीकरण का आशय यह देखना होता है कि हम किसी प्रलेख के सारांश अथवा शीर्षक में विद्यमान पदों पर भरोसा करें। प्राकृतिक भाषा अनुक्रमणीकरण, किसी प्रलेख को कैसे अभिलिखित किया गया है इसके आधार पर, उस प्रलेख की संपूर्ण विषय-वस्तु पर आधारित होता है। संभवतया यह प्रक्रिया प्रलेख को बहुत ही विस्तारित अनुक्रमणीकरण की ओर ले जाये अथवा किसी विशिष्ट प्रलेख संबंधों के संदर्भ में कौन से पद अत्यन्त महत्वपूर्ण हैं, यह निर्णय लेने के लिए इसमें कोई विधि स्थापित हो जाये। कम्प्यूटरीकृत अनुक्रमणीकरण में इस प्रक्रिया में, पदों के वर्तमान होने के सापेक्षिक आवर्तन का सांख्यीकीय विश्लेषण शामिल होगा। मानवीय रूप में अनुक्रमणी तैयार करने में इन पदों के चयन में कुछ निर्णय लेना अपेक्षित होगा। इसमें से अनेक समस्याओं को अनुक्रमणीकरण को शीर्षक और साराशों तक सीमित रख कर कम किया जा सकता है। कम्प्यूटर अथवा व्यक्ति, दोनों के द्वारा प्राकृतिक भाषा अनुक्रमणीकरण, कार्यान्वित किया जा सकता है। कम्प्यूटर अनुक्रमणीकरण में कम्प्यूटर अनुक्रमणिका के लिए उपयोगी समझे जाने वाले पदों की सूची का भली-भांति प्रयोग किया जा सकता है (उदाहरण एक प्रकार का समान्तर शब्द कोश), ताकि उचित पदों को पहचाना जा सके। प्राकृतिक भाषा के प्रयोग को चित्र 15.1 में दर्शाया गया है।



चित्र 15.1 प्राकृतिक भाषा सूचीपन

टिप्पणी

15.7.2 अनुक्रमणीकरण की मुक्त भाषा

अनुक्रमणिका की मुक्त भाषा पदों (टर्म्स) की सूचीबद्ध भाषा नहीं है, और विषय क्षेत्र में अवधारणाओं का वर्णन करने के लिए प्रयुक्त किये जाने वाले पदों से भिन्न हैं। अनुक्रमणिका इस अर्थ में 'मुक्त' है कि अनुक्रमणिका की प्रक्रिया में उन पदों (टर्मस) पर कोई प्रतिबंध नहीं हैं, जिनको इस प्रक्रिया में प्रयुक्त किया जा सकता हैं यहाँ तक कि मुक्त भाषा अनुक्रमणिका में इन प्रतिबंधों की कोई मान्यता ही नहीं है। मुक्त भाषा अनुक्रमणिका मानवों अथवा कम्प्यूटरों, दोनों के द्वारा किया जा सकता है। जब इसका कार्यान्वयन, विषय और इसकी शब्दावली के गहन अध्ययन सहित, किसी मानव द्वारा किया जाता है तो मुक्त भाषा अनुक्रमणीकरण के परिणामस्वरूप ऐसी अनुक्रमणी तैयार हो सकती है जो अनुक्रमणी के पदों को निर्धारित करने; और उपयोक्ताओं के परिप्रेक्ष्य से उनकी अपेक्षाओं में दोनों के साथ संगत होगा। तथापि सफलतापूर्वक ''मुक्त भाषा अनुक्रमणिका एक अनुक्रमणीकार के कार्य कौशल पर निर्भर होगा। कम्प्यूटरीकृत मुक्त भाषा अनुक्रमणीकरण, सभी व्यावहारिक प्रयोजनो के लिए प्राकृतिक भाषा अनुक्रमणिका के जैसा ही होगा।

यह अनुक्रमणीकरण की मुक्त भाषा की प्रकृति है कि विषय के लिए समुचित समझे जाने वाले किसी शब्द अथवा शब्द पद को अनुक्रमणीकरण के पद के रूप में निर्धारित किया जा सकता है। यह पद मशीन अथवा मनुष्य किसी के भी द्वारा निर्धारित किए जा सकते है, यद्यपि मशीन द्वारा अनुक्रमणिका के परिवेश में मुक्त भाषा सर्वाधिक लोकप्रिय है/कम्प्यूटर द्वारा अनुक्रमणिका में कम्प्यूटर के संचालन के दौरान इसे उपलब्ध कराये गये प्रत्येक शब्द का प्रयोग होता है, बशर्ते कि अन्यथा कोई अन्य विशेष अनुदेश इस संबंध में दिया गया हो।

नियंत्रित शब्दावलियाँ प्रायः मुक्त पाठ्य वस्तु की बिल्कुल सही खोज में सुधार लाती है, पुनः प्राप्ति सूची में से असंबद्ध पदों/मुहावरों को कम करती हैं। मुद्रित अनुक्रमणी बनाने और कम्प्यूटरीकृत डाटाबेस और डाटाबैंको तक पहुँच सुनिश्चित करने के लिए प्राकृतिक भाषा अनुक्रमणीकरण और कुछ सीमा तक मुक्त भाषा अनुक्रमणीकरण दोनों का प्रयोग अनुक्रमणी बनाने के लिए किया जाता है।

15.7.3 नियंत्रित अनुक्रमणिका भाषा

किसी भी पुन प्राप्ति पद्धति की शब्दावली (अनुक्रमणीकरण भाषा) का इसके कार्य निष्पादन पर काफी गहरा प्रभाव पड़ता है। यह शब्दावली खोज रणनीति बनाने और वास्तविक खोज करने (मिलान और डाटाबेस) के कार्य को प्रभावित करती है अनुक्रमणिका की नियंत्रित भाषायें वे भाषाायें है जिनमें विषयों का संदर्भ देने के लिए जिन पदों का प्रयोग किया जाता है उन पदों और जिस प्रक्रिया के द्वारा किसी विशिष्ट प्रलेख के लिए कौन से पद का संदर्भ दिया जाना है उस प्रक्रिया का निर्धारण किसी व्यक्ति द्वारा नियंत्रित अथवा कार्यान्वित किया जाता है। साधारणतया, पदावलियों की एक सूची उपलब्ध रहती है जो प्रलेखों को पहचानने के लिए निर्धारित किए जाने वाले संभावित पदों की प्राधिकृत सूची के रूप में कार्य करती है। अनुक्रमणी बनाने के इस कार्य में, एक व्यक्ति सम्मिलित होता है जो विशिष्ट प्रलेखों की पहचान के लिए इस सूची में से पदों को निर्धारित करता है। दो प्रकार की नियंत्रित अनुक्रमणीकरण भाषायें होती

हैं: वर्णानुक्रमानुसार अनुक्रमणीकरण भाषायें और वर्गीकरण पद्धतियाँ वर्णानुक्रमानुसार अनुक्रमणीकरण भाषाओं में, जैसे कि समांतर कोश और विषय शीर्ष, विषयों के पद, विषयों के नामों की वर्णानुक्रमानुसार बनाई गई सूची होती है। जिन पदों का प्रयोग होता है, वे पद विषय के नाम के लिए पूर्वनिर्धारित होते हैं। परन्तु अन्यथा यह पद सामान्य शब्द ही होते हैं। वर्गीकरण पद्धतियों में, प्रत्येक विषय के लिए संकेत लिपि में एक शब्द चिन्ह नियत किया जाता है। आमतौर पर संकेत लिपि निर्धारित करने का प्रयोजन यही होता है कि अन्य विषयों के परिप्रेक्षय में उस विषय को उन विषयों की संदर्भ सीमा के अंदर ही रखा जा सके। वर्गीकरण पद्धतियों और वर्णानुक्रमानुसार अनुक्रमणीकरण भाषायें, दोनों को ही अनेक प्रकार के संदर्भों के लिए प्रयोग में लाया जाता है। दोनों प्रकार की युक्तियों का अनुप्रयोग आपको इनमें देखने को मिल सकता है, जैसे प्रसूचियाँ पुस्तकों और सामयिक पत्रिकाओं के अनुक्रमणी, संदर्भ ग्रंथ-सूचियाँ, वर्तमान के प्रति जागरूकता संबंधी बुलेटिन, वितरण के लिए चयनित विशिष्ट सूचना, कम्प्यूटरीकृत डाटाबेस, और डाटाबैंक, सारांश और अनुक्रमणीकरण सेवाएँ, विश्वकोश शब्दकोश और मार्ग-निर्देशिकाएँ। प्रलेखों के भौतिक व्यवस्थापन में वर्गीकरण का विशिष्ट अत्यन्त महत्त्व है।

साधारणतया पदों की कोई सूची, विषय शीर्षकों की सूची अथवा समांतर कोश उपलब्ध होता है, जो प्रलेखों की पहचान के लिए उनके लिए पदों का निर्धारण करने की प्रक्रिया में, प्राधिकृत सूची के रूप में कार्य करता है। अनुक्रमणीकरण में, इस सूची में से विशिष्ट प्रलेखों के लिए पद निर्धारित करने का कार्य शामिल है। खोजकर्ता से यह अपेक्षा की जाती है कि वह खोज की रणनीति तैयार करने के दौरान इसी नियंत्रित सूची से परामर्श ले। अत: प्रलेख का वर्णन करने के लिए अनुक्रमणीकार द्वारा प्रयुक्त किए जाने के लिए यही एकमात्र शब्द समूहों/पदों की उपलब्ध सूची है। अत: पाठ्य वस्तु मुक्त खोज की तुलना में नियंत्रित शब्दावली का प्रयोग सूचना पुन: प्राप्ति प्रणाली के कार्य निष्पादन में संवृद्धि करता है।

पाठगत प्रश्न 15.5

1. ''अनुक्रमणीकरण प्राकृतिक भाषा'' पद की व्याख्या करें।
2. ''मुक्त अनुक्रमणीकरण भाषा' की प्रकृति क्या है?
3. नियंत्रित अनुक्रमणीकरण भाषा का वर्णन करें।

आपने क्या सीखा

- ''सूचना पुन: प्राप्ति पद का निर्माण, 1950 में कैल्विन मूर द्वारा किया गया था।
- 'सूचना पुन: प्राप्ति', सूचना संसाधनो के संग्रह में से अपेक्षित सूचना के संदर्भ में, सूचना संसाधनों को प्राप्त करने की गतिविधि है।
- सूचना का संग्रहण और पुन: प्राप्ति, सूचना का व्यवस्थापन और पुन: प्राप्ति, पाठ्य-वस्तु

की पुनः प्राप्ति, सूचना का प्रस्तुतिकरण और पुनः प्राप्ति और सूचना तक पहुँच, सूचना पुनः प्राप्ति के विभिन्न लक्ष्यार्थ है।

- कोई भी ग्रंथालय, अपने संग्रह के प्रलेखों में से सूचना खोजने के लिए, सूचना पुनः प्राप्ति के कार्य को पूरा करने के लिए किसी न किसी प्रणाली को अपनाता है।
- आधुनिक सूचना पुनः प्राप्ति प्रणाली का संबंध, सूचना के भंडारण, व्यवस्थापन और विषय-वस्तु तक पहुँच के साथ-साथ मल्टीमीडिया सूचना संसाधनों के साथ भी है।
- आई आर एस का मुख्य उद्देश्य है सूचना की पुनः प्राप्ति फिर चाहे वह वास्तविक सूचना हो अथवा उन प्रलेखों के माध्यम से जिनमें वह सूचना सम्निहित हो चाहे वह स्थानापन्न सूचना हो, जो उपयोक्ता की प्रश्नावली से पूर्णतया अथवा आंशिक रूप से मेल खाती हो।
- विषय के अनुसार विस्तृत व्यवस्थापन पर विचारण करने वाले सर्वप्रथम ग्रंथालयी थे मेल्विल ड्यूई।
- अनुक्रमणीकरण की प्राकृतिक भाषायें वास्तव में कोई पृथक भाषा नहीं होती परन्तु यह जिस प्रलेख की अनुक्रमणी बनायी जा रही होगी उसकी 'प्राकृतिक भाषा' अथवा सामान्य भाषा होती है।
- अनुक्रमणीकरण की मुक्त भाषायें उन पदों की सूचीबद्ध भाषा नहीं हैं जो किसी प्रलेख में विषय का वर्णन करने के लिए प्रयोग किए जाने वाले पदों से भिन्न हो, वरन अनुक्रमणीकार द्वारा प्रदत्त प्रलेख में वर्णित विषय-वस्तु के योग्य पद होते हैं।
- नियंत्रित अनुक्रमणीकरण भाषा, अनुक्रमणीकरण की वे भाषायें हैं जिसमें विषयों के प्रतिनिधित्व के लिए प्रयोग किए जाने वाले पद और वह प्रक्रिया जिसके द्वारा किसी विशिष्ट प्रलेख के लिए पद निर्धारित किये जाते हैं, इन दोनों को किसी व्यक्ति द्वारा नियंत्रित अथवा कार्यन्वित किया जाता है।

पाठांत प्रश्न

1. एक सूचना पुनः प्राप्ति प्रणाली के उद्देश्य क्या हैं?
2. सूचना पुनः प्राप्ति प्रणाली के प्रमुख कार्यों की विवेचना करें।
3. अनुक्रमणीकरण की प्राकृतिक, मुक्त और नियंत्रित भाषाओं में अंतर स्पष्ट करें।
4. चित्र की सहायता से प्राकृतिक भाषा के प्रयोग की व्याख्या करें।

पाठगत प्रश्नों के उत्तर

15.1

1. सूचना पुनः प्राप्ति, सूचना संसाधनों के भंडार में से अपेक्षित सूचना की अपेक्षा के संदर्भ

में सूचना संसाधनों को प्राप्त करने की गतिविधि है। यह ग्रंथालय के अत्यंत महत्वपूर्ण कार्यों में एक है, क्योंकि यह कार्य किसी उपयोक्ता द्वारा अपेक्षित सूचना की मांग को पूरा करती है।

15.2

1. किसी भी सूचना पुन: प्राप्ति प्रणाली का मुख्य उद्देश्य है सूचना की पुन: प्राप्ति करना। यह वास्तविक सूचना भी हो सकती है अथवा उन प्रलेखों के माध्यम से भी जिनमें पूर्णतया अथवा आंशिक रूप से उस स्थानापन्न सूचना का समावेश हो, जो की प्रश्नावली से मेल खाती है।

15.3

1. ग्रंथालय प्रसूची एक साधन है जो ग्रंथालय में उपलब्ध प्रलेखों की-अवस्थिति को दर्शाता है। यह प्रसूची प्रलेख में निहित सूचनाप्रदान नहीं करती है, उदाहरण रूप में पत्रिका में प्रकाशित लेख, आदि। अनुक्रमणी, इस सूचना को ग्रंथात्मक सारांश और अन्य ग्रंथात्मक साधनों के द्वारा प्रदान की जाती है।

15.4

1. विषय के रूप विस्तृत व्यवस्थापन के संबंध में सर्वप्रथम विचार करने वाले ग्रंथालयी थे मेल्विल डयूई। ड्यूई से पहले भी ग्रंथालयी अपने ग्रंथालयों की व्यवस्था वर्गीकृत रूप में ही करते होंगे; क्योंकि वर्गीकृत प्रसूची काफी लोकप्रिय थी। तथापि उनकी वर्गीकृत व्यवस्थापन मोटे तौर पर विषय समूहों में था, विषय का विस्तृत विशिष्टकरण करने का कोई प्रयास नहीं किया गया था जिसे ड्यूई ने पहली बार सुझाया था कि यह अनिवार्य और उपयोगी है।

15.5

1. 'प्राकृतिक भाषा में अनुक्रमणीकरण में विषय वस्तु को दर्शाने के लिए पद उसी प्रलेख से लिए जाते हैं।
2. अनुक्रमणीकरण की मुक्त भाषा की प्रकृति है कि कोई भी शब्द अथवा पद, जो विषय के लिए उपयुक्त हो उसे अनुक्रमणीकरण पद के रूप में निर्धारित किया जा सकता है।
3. अनुक्रमणीकरण की नियंत्रित भाषायें वे भाषायें हैं जिनमें वे पद जिन्हें विषयों का प्रतिनिधित्व करने और वह प्रक्रिया जिसके माध्यम से किसी विशिष्ट प्रलेखों के लिए पदों, जिन्हें किसी व्यक्ति द्वारा नियंत्रित अथवा कार्यान्वित किया जाता हो।

पारिभाषित शब्दावली

आंकड़ों की पुन: प्राप्ति (Data Retrieval) : उस सूचना की पुन: प्राप्ति जिसमें निहित विषय-वस्तु उपयोक्ता द्वारा अपेक्षित प्रश्नावली की सूचना आवश्यकता को संतुष्ट करती हो।

टिप्पणी

अनुक्रमणी पद (Index Term) : कोई भी पूर्व-चयनित ऐसा पद, जिसे किसी प्रलेख की विषय-वस्तु का संदर्भ देने के लिए प्रयोग किया जाता है।

सूचना पुनः प्राप्ति (आईआर) (Information Retrieval (IR) : एक वृहद् भंडार (प्रायः कम्प्यूटरों में संग्रहित) में से वह सामग्री (सामान्यतया प्रलेख ढूंढना जो किसी सूचना की मांग को संतुष्ट करती हो।

मुख्य शब्द (Keyword) : अनुक्रमणी पद के समान।

प्रश्नावली (Query) : उपयोक्ता की सूचना आवश्यकता की अभिव्यक्ति

पुनः प्राप्ति (Retrieval): किसी उपयोक्ता के अनुरोध के प्रत्युत्तर में सूचना प्रणाली द्वारा निष्पादित कार्य।

उपयोक्ता की सूचना अपेक्षा (User Information Need) : किसी उपयोक्ता की सूचना अपेक्षा का प्राकृतिक भाषा में कथन।

शब्दावली (Vocabulary) : पाठय वस्तु में निहित सभी शब्दों का समुच्चय (सेट)

वेबसाइट्स

http://en. wikipedia.org/wiki/information.retrieval

http://polaris.goels.uda.edu/page/is227.html

http://nlp.stanford.edu/IR-book