એકમ–4 ડેટાબેઝ સર્ચિંગ DATABASE SEARCHING

: માળખું :

4.0 ઉદ્દેશ
4.1 પ્રસ્તાવના
4.2 ઈન્ફોર્મેશન રીટ્રાઈવલ
4.3 ઈન્ફોર્મેશન રીટ્રાઈવલ વિરુદ્ધ ડેટા રીટ્રાઈવલ
4.4 સર્ચઆઉટપુટના ઉદ્ભવ માટે પેરામીટર્સ - (માનાંક)
4.5 સર્ચ સ્ટ્રેટેજી
4.6 સંયુક્ત ક્વેરીઝ
4.7 પ્રગતિશીલ ફીચર્સ
4.8 Keyword Grouping
4.9 Truncation
4.10 Proximity Operators
4.11 Limiting Search
4.12 Trends in Information Retrieval
4.13 સારાંશ
4.14 તમારી પ્રગતિ ચકાસોના ઉત્તરો
4.15 ચાવીરૂપ શબ્દો
4.16 સંદર્ભ અને વિશેષ વાંચન

4.0 ઉદ્દેશ (OBJECTIVES)

- **આ પ્રકરણ પૂરું થયા પછી તમે નીચે પ્રમાણે જાણી શકશો.**
  - સર્ચ પ્રોસેસ અને અસરકારક સર્ચ કરવા માટેના સ્ટેપ્સ.
  - સર્ચ લેંગવેજ માટેના લાક્ષણિક લક્ષણો જેમાં વિવિધ સર્ચ ટેક્નિક અને સર્ચને આગળ વધારવા માટેનાં સાધનોનો પણ સમાવેશ થાય.

અહીં ડેટાબેઝ સર્ચીંગનો ઉપયોગ આનુષંગિક માહિતીની પુનઃ પ્રાપ્તી કરવા માટે કરવામાં આવે છે. દા.ત. ઈન્ટરનેટ / વેબ સર્વીસીઝ, ઓન લાઈન ડેટાબેઝ સર્ચીસ [દા.ત. ઓન લાઈન ડેટાબેઝીસ ડેટાબેઝ દ્વારા પ્રદાન કરે છે. I. e., DIALOG, BRS,વગેરે] OPAC નું સર્ચીંગ આ સમગ્ર વિવરણ કે જેને 'સર્ચ એન્જિન' અને 'સર્ચ લેંગવેઝ' કહેવામાં આવે છે. તેનો ઉપયોગ વારાફરતી કરવામાં આવે છે.

4.1 પ્રસ્તાવના (INTRODUCTION)

આ પ્રકરણના ઉપયોગ માટે આપણે ડેટાને રો ડેટા તરીકે ઓળખીશું. નંબર અથવા આંક કે જે હકીકતનું બયાન કરે છે અને કોમ્પ્યુટરમાં સ્ટોર કરે છે. કોમ્પ્યુટરમાં આવેલા સ્ટોર કરેલા ડેટાનું ફિઝીકલ ફોર્મમાં અથવા લોજીકલ ફોર્મમાં મહત્ત્વ નજીવું છે કારણ કે કોઈપણ હકીકત કોઈપણ સંદર્ભ વગર ઉપયોગી નથી. ડેટાબેઝ એ ડેટાનો સ્ટોર ડેટાનો સંગ્રહ છે કે જેનું અસ્તિત્વ માહિતી પૂરી પાડવા માટે છે. જો કે ડેટાબેઝ મેન્યુઅલ અથવા કોમ્પ્યુટરાઈઝડ પણ હોઈ શકે. આ પ્રકરણના ઉપયોગ માટે ડેટાબેઝ એ કોમ્પ્યુટર ફાઈલમાં આવેલા ડેટાનો સંગ્રહ છે, કે જે સામાન્ય રીતે મલ્ટીયુઝર્સ માટે પ્રાપ્ય છે. હાલના તબક્કે ડેટાબેઝને ડેટાબેઝ સિસ્ટમથી કે જેનાં બધાં ઘટકો એક બીજાને એકત્ર કરે છે. મેનીપ્યુલેટ, મેનેજસર્ચ, રીટ્રાઈપલ અને માહિતીને પહોંચાડવાનું કાર્ય કરે છે. તેને જુદુ પાડવું અગત્યનું છે. સામાન્ય રીતે ઉપયોગમાં આવતુ ડેટાબેઝ મેનેજમેન્ટ સિસ્ટમ (DBMS)ડેટાબેઝ સિસ્ટમની એક જાતનું છે, કે

જે હાઈલી સ્ટ્રક્ચર્ડ ડેટાની પ્રક્રિયા માટે રચવામાં આવેલું છે. માહિતીની પુનઃ પ્રાપ્તીએ કાયમ માટે બહુ જ આદર્શ ઉલ્લેખ કરતી ટર્મ છે. અને આપણે તેને મોટેભાગે ઓટોમેટેડ ઈન્ફોર્મેશન રીટ્રાઈવલ સિસ્ટમ માટે ઉપયોગ કરીએ છીએ. અરે, ઈન્ફોર્મેશન શબ્દ પણ ગેરમાર્ગે દોરે છે. ઈન્ફોર્મેશન રીટ્રાઈવલ (IR) (SHANNON) એ ઉપયોગ કર્યો તે પ્રમાણે ટેક્નિકલ દૃષ્ટિએ ઈન્ફોર્મેશનએ સહજતાથી માપવામાં આવ્યું નથી. તે કદાચ ઘણી પરિસ્થિતિમાં માત્ર ડોક્યુમેન્ટને 'ઈન્ફોર્મેશન'ની અવેજીમાં મૂકવાથી પુનઃ પ્રાપ્તિની સ્થિતિને યોગ્ય રીતે વર્ણન કરવા માટે વધારે યોગ્ય ગણાશે. આમ છતાં 'ઈન્ફોર્મેશન રીટ્રાઈવલ' ટેક્નિકલી ટર્મ તરીકે વ્યાપક પ્રમાણમાં સ્વીકૃતિ ધરાવે છે. અને સામાન્ય રીતે તેનો ઉપયોગ સાહિત્યમાં Cleverdon, Salton, Sparckjones, Lancaster અને અન્યની રચનાઓમાં વર્ણવેલી પ્રવૃત્તિનું વર્ણન કરવામાં આવે છે. 'એન ઈન્ફોર્મેશન રીટ્રાઈવલ સિસ્ટમ' (Text Retrievel System) તરીકે પણ ઓળખાય છે. ટેકસટ્યુઅલ રેકોર્ડમાંથી માહિતી પુનઃ પ્રાપ્ત કરવા માટે, સિસ્ટમ પ્રક્રિયાને રચવા માટે યોજે છે. મોટાભાગની ઈન્ફોર્મેશનરીટ્રાઈવલ સિસ્ટમ ફક્ત વપરાશકર્તાને ઉપસ્થિતિની (અથવા Non - Existence) તેણે વિનંતી કરેલા ડોક્યુમેન્ટના વ્હેર એબાઉટની માહિતી આપે છે. આપણે બધાં જાણીએ છીએ કે ઈન્ટરનલ ઓર્ગેનાઈઝેશન મોટાભાગની ટેક્સ્ટ્યુઅલ એન્ટીટી સ્ટ્રકચર - ટુ - લેસ સ્ટ્રકચર બદલાતું રહે છે અને ક્યારેક અન્સ્ટ્રકચર્ડ, સામાન્ય રીતે ડેટાબેઝમાં ડેટાનું નાનામાં નાનુ અર્થપૂર્ણ યુનિટ એ રેકોર્ડ છે અને તેનો મેળ બેસે છે. એન્ટીટી / thing સાથે - વાસ્તવિક વિશ્વમાં ઈન્ફોર્મેશન રીટ્રાઈવલ સિસ્ટમમાં રેકોર્ડ એ ટેક્સ્ટ્યુઅલ એન્ટીટીનું અભિન્ન અંગ છે. દા.ત. પુસ્તકો, વર્તમાનપત્રો, સામયિકો વગેરે ડેટાબેઝ સિસ્ટમના જુદી જુદી જાતના ડેટાબેઝના ડેટાના વિવિધ પ્રકારની અને ડેટાની પ્રક્રિયાની જરૂરિયાતને ધ્યાનમાં રાખી આર્કિટેકચરે વિકાસ કર્યો છે. આ યુનિટમાં ઈન્ફોર્મેશન રીટ્રાઈવલ સિસ્ટમ (IRS) પર ફોક્સ કરવામાં આવ્યું છે. I. e.મેળવવા માટે અને ટેક્સટ્યુઅલ ડેટાબેઝને સર્ચ કરવા માટે બિબ્લિઓગ્રાફિક ડેટાબેઝ પર ખાસ ભાર મૂકવામાં આવ્યો છે.

4.2 ઈન્ફોર્મેશન રીટ્રાઈવલ (INFORMATION RETRIEVAL) :

ઈન્ફોર્મેશન રીટ્રાઈવલ એ શબ્દ (ટર્મ) પાંચ દાયકા પહેલાં CALVIN MOOERS એ બધી જ પ્રક્રિયાને સંગઠીત કરી, જે ડેટાબેઝના ડોક્યુમેન્ટ અને આનુસંગીક ક્વેરી માટે સર્ચીંગ અને રીટ્રાઈવીંગ ઈન્ફોર્મેશન સાથે સંકળાયેલો છે. IR સિસ્ટમની મૂળભૂત પ્રક્રિયાને ડાયાગ્રામ દ્વારા આકૃતિ 4માં દર્શાવી છે.

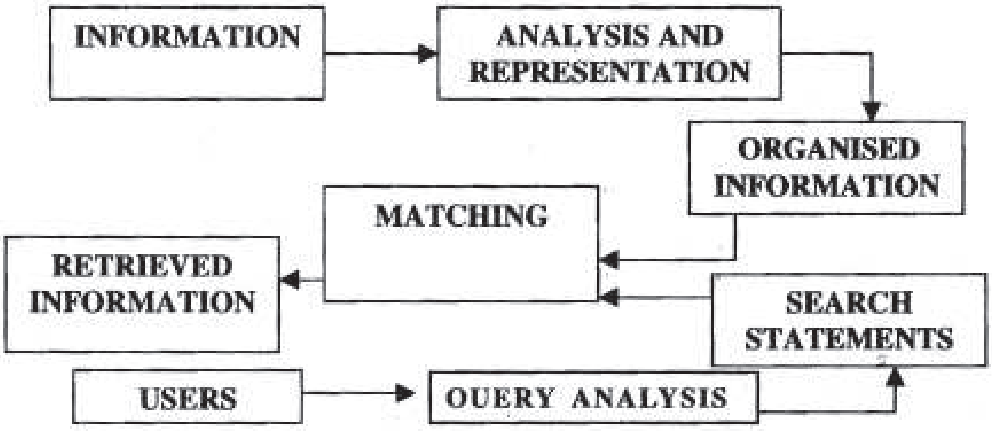

Fig. 4.1: Process of Information Retrieval

(Source : Chowdhury, G.G. Introduction to Modern Information Retrieval. – London: Library Association, 1999. – p.4)

આકૃતિ

ડાયાગ્રામ (IRS)ની મુખ્ય કામગીરીને પ્રતિબિંબિત (High Light) કરે છે. દા.ત. રીટ્રાઈવલ અને માહિતીની જરૂરિયાતને અનુલક્ષીને સર્ચ અને માહિતીની પુનઃ પ્રાપ્તિ માટે માહિતીનું સંગઠન કરે છે. રીટ્રાઈવલ પ્રક્રિયા સામાન્ય રીતે વપરાશકર્તાની માહિતી માટેની ક્વેરીથી શરૂ થાય છે. સર્ચ શબ્દ એ

ઈન્ડેક્ષ ફાઈલ સાથે અનુરૂપ છે. (અથવા ડેટાબેઝમાં ટર્મ સાથે) અને અનુરૂપ રેકોર્ડ પુનઃ પ્રાપ્ત થાય છે. જ્યારે ઈન્ફોર્મેશન રીટ્રાઈવલ માટેની બધી જ કાર્યવાહીમાં આ મૂળભૂત પ્રક્રિયા છે. વાસ્તવમાં IR સિસ્ટમ અને તેથી રીટ્રાઈવલની પ્રક્રિયા સ્વભાવમાં જટિલ છે. અને સંતોષકારક અને સ્વીકાર્ય પરિણામ પ્રાપ્ત થાય તે પહેલાં ઘણાં ઓપરેશન અને સાધનોની જરૂર પડે છે.

◆ **તમારી પ્રગતિ ચકાસો**

1. ઈન્ફોર્મેશન રીટ્રાઈવલ એટલે શું ? ઈન્ફોર્મેશન રીટ્રાઈવલમાં સંકળાયેલી વિવિધ પ્રક્રિયાને (ઓળખાણ આપી જણાવો) આઈડેન્ટીફાય કરો ?

નોંધ (i) તમારા જવાબો નીચે આપેલ જગ્યામાં લખો.

(ii) આ પ્રકરણના અંતે આપેલા જવાબો સાથે તમારા ઉત્તરો સરખાવો

....

....

....

....

....

....

4.3 ઈન્ફોર્મેશન રીટ્રાઈવલ વિરુદ્ધ ડેટા રીટ્રાઈવલ (INFORMATION RETRIEVAL VERSUS DATA RETRIEVAL)

ડેટાબેઝ સર્ચીંગની વિગતોમાં ઉતર્યા પહેલાં, 'ઈન્ફોર્મેશન રીટ્રાઈવલ' અને ડેટા રીટ્રાઈવલ વચ્ચેનો તફાવત સમજવો ઉપયોગી રહેશે. જો કે ઘણાં ટીકા કરે છે કે બે વચ્ચેની સહરદ બહુ સ્પષ્ટ નથી અને તે આધારે બે ભાગમાં વહેંચાયેલા છે. એ ખરેખર સાચુ છે કે બે વચ્ચેનો તફાવત અસ્પષ્ટ છે. આમ છતાં આપણે રીટ્રાઈવલની વાત કરીએ ત્યારે આપણે સામાન્ય રીતે બરોબર અનુરૂપની શોધ કરીએ છીએ. જ્યારે ઈન્ફોર્મેશન રીટ્રાઈવલમાં આ ક્યારેક રસનો વિષય હોઈ શકે પણ ઘણું ખરું આપણે શોધ કરીએ છીએ કે વિનંતી કરવામાં આવેલ આઈટમ, કે જે સાધારણ અનુરૂપ હોય અને તે પછી થોડુંક અનુરૂપ હોય તેની પસંદગી કરીએ છીએ. ઈન્ફોર્મેશન રીટ્રાઈવલમાં એ વાત સામાન્ય છે કે તેનો ઘણો ખરો ઉપયોગ ઈન્ડકટીવ ઈનફરન્સ રીલેશન, ચોક્કસતા અને અચોક્કસતાની માત્રાને સ્પષ્ટ કરે છે. અને તેથી ઈન્ફરન્સ (તર્કમાં) વિશ્વાસ એ અસ્થિર છે. તેથી આપણે ડેટા રીટ્રાઈવલને ડિટરમિનિસ્ટિક (દૃઢ નિશ્ચયી) જ્યારે ઈન્ફોર્મેશન રીટ્રાઈવલને પ્રોબેબલિસ્ટિક (સંભવિત) તરીકે વર્ણવીએ છીએ. તેથી શક્યતઃ આ બાબતમાં જુદા જુદા પ્રકારના સોફટવેરનો- ઈન્ફોર્મેશન રીટ્રાઈવલ સિસ્ટમ અને ટેક્સ્ટ રીટ્રાઈવલ સિસ્ટમનો વિકાસ કરવામાં આવ્યો છે અને તેથી આ પરિસ્થિતિમાં ઈન્ફોર્મેશન રીટ્રાઈવલ અને ડેટા રીટ્રાઈવલ વચ્ચેનો મુખ્ય તફાવત જાણવા માટે સામાન્ય વિચાર ઉપયોગ થશે. (IRS) ડોક્યુમેન્ટ અથવા તેના સરોગેટ કે જેમાંના ઘટકો લંબાઈ અને કેરેક્ટર (લક્ષણમાં) બદલાતા રહે છે તેના સ્ટોર કરે છે. ડોક્યુમેન્ટ એ સામાન્ય રીતે કોઈક પ્રકારના ટેક્ષ્યુઅલ રેકોર્ડ સંપૂર્ણ અને સ્ટ્રક્ચર્ડથી વધારે સ્ટ્રક્ચર્ડના ક્ષેત્રમાં આવે છે. આ એક મુખ્ય વિશેષતા છે કે જે (IRS) ડિઝાઈનીંગમાં ધ્યાન રાખવું પડે દા.ત. WINISIS, સોફટવેર પેકેજ, ટેક્સ્ટ રીટ્રાઈવલ સીસ્ટમ માટે લાઈબ્રેરી સંગ્રહ, રીસર્ચ પ્રોજેક્ટનો હોસ્પિટલ રેકોર્ડ ડેટાબેઝ વગેરેનો ઉપયોગ કરવામાં આવેછે. ડેટા રીટ્રાઈવલમાં આપણે સિસ્ટમ કે જે હાઈલી સ્ટ્રક્ચર્ડ ડેટાને મેનિપ્યુલેટ કરવા ડિઝાઈન કરવામાં આવ્યા હોય તે તરફ દૃષ્ટિ કરીએ છીએ. (IRS) રીટ્રાઈવલ ટેક્સ્યુઅલ રેકોર્ડને લવચિકતા આપવા માટે ખાસ કરીને વિકસાવવામાં આવ્યું છે. સર્ચની સગવડ જે (IRS) આપે છે અને દલીલ કરવામાં આવે છે કે તે ડેટાબેઝમાં સર્ચ ટર્મ અને ફિલ્ડ વેલ્યુ રેકોર્ડ વચ્ચે ચોક્કસ અનુરૂપ છે તે પૂરતી નથી. તેથી (IRS) અનુસંધાને ડેટાબેઝ સર્ચિંગ માગી લે છે કે રીટ્રાઈવલ રેકોર્ડને ટેકો આપવા ક્વેરી સાથે પ્રીડિફાઈન્ડ ડીગ્રી ઓફ સિમિલારિટી પૂરતા પ્રમાણમાં ધરાવે છે. બીજી બાજુ 'ડેટા રીટ્રાઈવલ'માં આપણે સંપૂર્ણ અનુરૂપતા શોધીએ છીએ.

◆ **તમારી પ્રગતિ ચકાસો**

2. ઈન્ફોર્મેશન રીટ્રાઈવલ અને ડેટા રીટ્રાઈવલ વચ્ચેના તફાવતની મુખ્ય લાક્ષણિકતાઓ સ્પષ્ટ કરો.

નોંધ (i) તમારા જવાબો નીચે આપેલી જગ્યામાં લખો

(ii) તમારા જવાબો આ પ્રકરણના અંતે આપેલા ઉત્તરો સાથે સરખાવો.

....

....

....

....

....

....

4.4 સર્ચઆઉટપુટના ઉદ્‌ભવ માટે પેરામીટર્સ - (માનાંક) (PARAMETERS FOR EVALUTATION OF SEARCH OUTPUT)

વ્યાપક રીતે ઉપયોગમાં આવતા બે મુખ્ય માનાંક IRS માં સર્ચઆઉટપુટને ચકાસવા માટે ઉપયોગમાં આવે છે તે છે રી-કોલ - IRSની સંબંધિત ડોક્યુમેન્ટને પુનઃ પ્રાપ્ત કરવાની શક્તિનું માપ અને પ્રિસિજન : (IRS)ની રીટ્રાઈવલમાંથી બિનજરૂરી ડોક્યુમેન્ટને પરત કરવાની (સંક્ષેપ) શક્તિનું માપ.

વપરાશકર્તાની જરૂરિયાતની રીકોલ અને પ્રીસીઝનની ઈચ્છિત કક્ષા ઘણી બાબતો ઉપર આધાર રાખે છે. IRSની સર્ચ લેંગ્વેજ, ઘણુંખરું રીકોલ અને પ્રીસીઝનને નિયમસર કરવા માટે વ્યૂહરચનાને સુધારાવધારા માટે ટેકો આપે છે. સાહિત્યમાં ત્રણ જુદી જુદી જાતના સર્ચને દર્શાવવામાં આવ્યા છે.

હાઈકોલ સર્ચ : વપરાશકર્તાઓ તેની ક્વેરી બાબત બધી જ સંબંધિત આઇટમને શોધવાની ઈચ્છા કરે છે.

હાઈપ્રિસીઝન સર્ચ : વપરાશકર્તા ફક્ત વિષયને લગતી આઇટમ શોધવાનું અને બને ત્યાં સુધી બિનજરૂરી આઇટમને નિવારણ માટે ઈચ્છા છે.

બ્રીફ સર્ચ : વપરાશકર્તા ફક્ત કેટલીક આનુષાંગિક આઇટમને શોધવાનું ઈચ્છે છે. (Medows and cochrane, Basics of Online Searching, 1981)

4.5 સર્ચ સ્ટ્રેટેજી (SEARCH STRATEGY)

કાર્યદક્ષ અને અસરકારક સર્ચ કરવા માટે વિકાસ અને યોગ્ય સર્ચ વ્યૂહરચનાનો સમાવેશ કરવાની જરૂર પડે છે. અહીં સર્ચ સ્ટ્રેટેજીનો અર્થ સર્ચ પ્રક્રિયામાં ઓપરેશન્સ અને નિર્ણયોની વપરાશકર્તાની જરૂરિયાતને નિરાકરણ ન આવે ત્યાં સુધી (IR સિસ્ટમની સર્ચ લેંગ્વેજની વાક્ય રચનાનો ઉપયોગ કરીને) પૂરતી અને ચોક્કસ માહિતીને વ્યક્ત કરે નહીં ત્યાં સુધી સિસ્ટમનો ઉપયોગ કરે છે. સ્પષ્ટ રીતે ઘણાં સ્ટેપ્સ અને મુદ્દાની સમજણ દરેક તબક્કે અસરકારક અને કાર્યદક્ષ સર્ચ માટે પૂર્વાશ્યક છે.

1. Understanding the User's Requirements (વપરાશકર્તાની જરૂરિયાતને સમજવી) :

વપરાશકર્તા માહિતી મેળવવાની જરૂરિયાતને સમજવા માટે સર્ચ પ્રક્રિયાને અને આઉટપૂટને અર્થપૂર્ણ રીતે અસર કરે છે. એ મહત્ત્વનું છે કે વપરાશકર્તાની જરૂરિયાત સર્ચ પ્રક્રિયાને શરુ કરતા પહેલાં જેટલી બને તેટલી સ્પષ્ટ હોવી જોઈએ. આથી સ્પષ્ટ યોજના નીચે પ્રમાણે મેળવી શકાય.

◆ **વપરાશકર્તાની મુલાકાત** : મેડીએટેડ સર્ચીસમાં આ પ્રક્રિયા માહિતીના માધ્યમ દ્વારા છેવાડેના વપરાશકર્તાની માહિતીની જરૂરિયાતને મેળવવા માટે સ્પષ્ટ સમજણ શોધે છે. સિમેન્ટીક ગેપ, ખરેખર જરૂરિયાત વ્યક્ત અને માહિતીની જરૂરિયાત (Actual need, expressed need as pravial by the intermation intermediary) વચ્ચે માહિતી માધ્યમ દ્વારા આ પ્રક્રિયા દરમ્યાન જોડાણ / સંકોચન કરવાનું કાર્ય શોધે છે. છેવાડાનો વપરાશકર્તા આ સર્ચ દરમ્યાન જો હાજર ન હોય તો આ ખાસ જરૂરી છે.

- ખરેખર સર્ચ કરતા પહેલા રીકોલની પસંદગીની લેવલને ધ્યાનમાં લેતા, જરૂરિયાતની સમજણ ચોક્કસ સહનશક્તિની પણ જરૂર પડે છે.

(B) સર્ચ કરવામાટે ડેટાબેઝની પસંદગી : એક વખતના વપરાશકર્તાની જરૂરિયાતને સમજ્યા પછી બીજું અગત્યનું પગથિયું છે ડેટાબેઝને પસંદ કરવાનું અથવા તો ડેટાબેઝને શોધવાનું. એ જાણવું જરૂરી છે કે યોગ્ય રેફરન્સને મેળવવાની શક્તિ ધરાવતા ઘણાં ડેટાબેઝીસ હોય છે. ડેટાબેઝની પસંદગીમાં અગત્યનું પાસુ તેના કાર્યક્ષેત્રમાં છે. ડેટાબેઝ, જો તે વિષય માટે / વપરાશકર્તાની જરૂરિયાતને યોગ્ય ન હોય તો અસરકારક નથી રહેતું. ડેટાબેઝ વ્યાપક રીતે તેના સ્કોપ/ આવરણ અને સમયોચિતતાના કારણે બદલાતા રહે છે. (Chowdhury) સૂચન કરે છે કે સર્ચ કરવાની જરૂર હોય તે ડેટાબેઝને પસંદ કરતી વખતે નીચેા મુદ્દાઓને ધ્યાનમાં રાખવાની જરૂર છે.

- Subject Coverage
- Document Coverage
- Accessibilty
- Period Coverage
- Search Fields
- Search Device/ Tools
- Prerformanc, etc.

(C) ફોર્મ્યુલેટિંગ ધ સર્ચ સ્ટેટમેન્ટ : ખરેખર સર્ચ કરતા પહેલા વિષય વિશે સ્પષ્ટ હોવું એ અગત્યનું છે. બધા જ શક્ય શબ્દો અથવા વાક્યો કે જેના ઉપયોગ વિષયને સ્પર્શતા હોય તેની જરૂરી ટર્મ (RTS) શબ્દના અંતે આપેલા શક્ય પર્યાઓ (દા.ત. એકવચન, બહુવચન, વિશેષણ) સમાનાર્થી શબ્દો, ભિન્ન ટર્મિનોલોજીના પર્યાયોનો વિચાર કરવો જોઈએ આ પછીના પેરેગ્રાફમાં ક્વેરી કેવી રીતે રચવી તે બતાવવામાં આવ્યું છે. ક્વેરી એ મૂળભૂત માહિતી છે કે જેસર્ચર (IR) સિસ્ટમને પહોંચાડે છે કે જેથી સિસ્ટમમાં માહિતી / માહિતીના સ્રોતને સ્પષ્ટ કરી શકાય જે જરૂરિયાત સાથે અનુરૂપ હોય, વાસ્તવિક રીતે તે ઘણાં પેરામીટર્સ (માનાંક) ધરાવે છે. જે એકત્ર થઈને જરૂરી માહિતી હકને સ્પષ્ટ કરેછે. આ પેરામીટર્સમાં સામાન્ય રીતે એક અથવા વધારે કી શબ્દો / વર્ણન કરતા, વર્ગીકરણ, સંકેત અને શક્યતઃ મૂળ ઘટક જે અન્ય પેરામીટર્સને વ્યાખ્યાયિત કરે છે. જેવા કે, તારીખ, ભાષા વગેરે ટર્મ કી-વર્ડઝનો અહીં ઉપયોગ કરવામાં આવશે. જેમાં કંટ્રોલ્ડ વોકેબ્યુલરી (વાક્ચાતુર્યનું નિયમન)માંથી પણ વર્ણનકર્તાનો સમાવેશ થશે. ક્વેરી સરળ હોવી જોઈએ છતાં જેની જરૂર છે તે પૂરતી વર્ણનાત્મક હોવી જોઈએ. તેમ ન થાય તો શોધકને પરિણામ ઉપર બહુ જ ઓછો અંકુશ રહેશે એની વધારે શક્યતા છે કે સર્ચ આઉટપુટ એ વિશાળ હોય, પ્રસ્તુત આઈટમને અનુરૂપ કરવા રીટ્રાઈવ્ડ ટેક્સ્ટની સ્ક્રીનીંગની આવશ્યકતા રહે. કમનસીબે એવી કોઈ એક ચોક્કસ રીત ક્વેરીને ઉકેલવા માટે નથી કે જે બધી જ પ્રસ્તુત માહિતી સિસ્ટમમાં પ્રાપ્ય હોય અને ક્વેરી એ મોટે ભાગે સર્ચ લેંગ્વેજ / સર્ચ એન્જિનનું કાર્ય હોય, જેનો ઉપયોગ કરવામાં આવ્યો હોય. આમ છતાં ચોક્કસ મૂળભૂત ફેક્ટર હોય છે, જે સમજવા માટે વધારે સારી ક્વેરીને રચવા માટે મદદરૂપ થશે. તેના સાદા ફોર્મમાં ક્વેરીનો સમાવેશ થાય છે - કી બોર્ડ અને / અથવા વાક્યોનો. જ્યારે વાક્ય (સાદા કી-વર્ડઝને સ્થાને) એ ઇનપુટ છે, ત્યારે ડોક્યુમેન્ટ કે જે વાક્ય ધરાવે છે તે પુનઃપ્રાપ્ત થશે. દા.ત. ડોક્યુમેન્ટ કે જેની પાસે વાક્ય છે 'રીટ્રાઈવલ ઓફ ઈન્ફોર્મેશન'એ વાક્યને અનુરૂપ થાય છે. આદર્શ રીતે સર્ચિંગમાં વપરાતું વાક્ય લાંબુ ન હોવું જોઈએ કારણ કે લાંબા વાક્યોને લીધે ચોક્કસ અનુરૂપ શોધવાની તક ઓછી થાય છે.

4.6 સંયુક્ત ક્વેરીઝ (COMPOUND QUERIES)

સર્ચના મૂળભૂત ઘટકો જેવા કે કી-વર્ડઝ અને વાક્યોને (અને અન્ય પેરામીટર્સ) વધારે શક્તિશાળી અને અસરકારક સર્ચીસ માટે જોડાણ કરી શકાય George Boole લોજીકલ ઓપરેટર્સનું એક સાધન શોધ્યું છે કે જે સાંકેતિક સ્ટેટમેન્ટને એક ફોર્મમાંથી બીજા ફોર્મમાં જટિલ સ્ટેટમેન્ટને ભેગા કરે છે. આવા ભેગા થયેલા દરેક બુલેન ઓપરેટર, ઈફેક્ટમાં રીલેશનશીપને વ્યક્ત કરે છે. John Veenn આવા

સંબંધોનું વર્ણન કરે છે અને આ પ્રમાણે (Boolan) ઓપરેટર્સ એ બહુ જ વ્યાપક રીતે ઉપયોગમાં આવતું સાધન કે જે હાલમાં સામાન્ય રીતે Venn Digrams તરીકે ઓળખાય છે. બુલીઅન ઓપરેટર્સ OR And અને NOT સેટ ઓપરેટર્સને Set Union, Set intersection અને Set disjunction ને સાદશ્ય કરે છે જેને ડાયાગ્રામની મદદથી નીચેના ઉદાહરણથી સમજાવવામાં આવ્યું છે.

Example Information
Information and Retrieval
Information Not Data

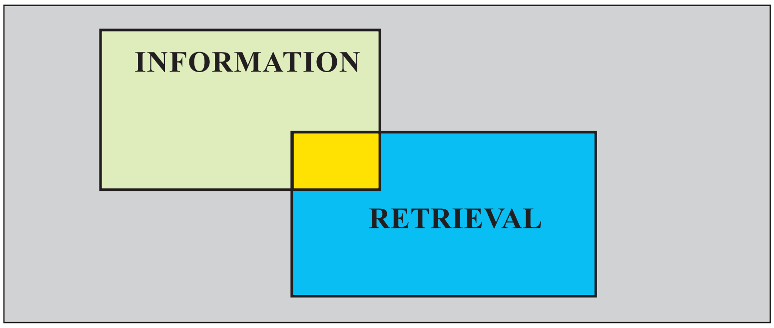

Fig. 4.2 : Database search using Boolean Operators

જ્યારે બે સર્ચ ટર્મ બુલીઅન ANDનો ઉપયોગ કરીને જોડવામાં આવે છે ત્યારે ફક્ત એ જ ડોક્યુમેન્ટ (પીળા કલરમાં બતાવેલા) પુનઃપ્રાપ્ત થશે. બે ટર્મને જોડવા માટે ORનો ઉપયોગ બન્ને લંબચોરસને સાદશ્ય કરતા ડોક્યુમેન્ટને પુનઃપ્રાપ્ત કરવામાં પરિણમશે. સર્ચ ઓપરેશન્સનો ઉપયોગ Infromation Not Retrival ફક્ત લીલા રંગમાં બનાવેલા વિસ્તારમાં આવેલા ડોક્યુમેન્ટને પુનઃ પ્રાપ્ત કરશે.

જ્યારે આ પ્રમાણે કોઈ એક અથવા બંને ટર્મ જોડાયેલા હોય ત્યારે OR ઓપરેટરનો ઉપયોગ કરવામાં આવે છે. બે અથવા વધારે પર્યાયો અથવા પર્યાયની નજીકમાં હોય ત્યારે OR નો ઉપયોગ વધારે યોગ્ય જણાશે. OR ઓપરેટર સર્ચને વિસ્તારે છે અને તેને એક સમાન માને છે. બીજી બાજુ AND અને NOT સર્ચને મર્યાદિત કરે છે. એક અને એ જ સર્ચ એક્સપ્રેશન બુલીઅન ઓપરેટરના એક કે વધારે મલ્ટીપલ ઉપયોગમાં સામેલ થાય છે. કોઈ પણ જાતના અપવાદ સિવાય બધી જ સર્ચ લેંગ્વેજીસ બુલીઅન ઓપરેટર્સનો સમાવેશ કરે છે. આનો ખરેખર અમલ કરવા માટે જુદા ઓપરેટર્સને પ્રતિનિધિત્વ માટે ઘણી સર્ચ સિસ્ટમ ચિહ્નનો ઉપયોગ કરે છે. (દા.ત. *)

ઉદાહરણ : (Information or Data and Retrieval[1])

બુલીઅન ઓપરેટર્સની થોડી ઘણી મોટી ખામીઓ છે

(a) બુલીઅન ઓપરેટર્સ રીટ્રાઈવલ ઓફ ડોક્યુમેન્ટ / Webpages કે જે કેટલીક ટર્મ અથવા ટર્મના સમૂહને માત્ર સ્વીકારે છે. વાસ્તવમાં રીટ્રાઈવલની જરૂરિયાતને બહુ જ સંક્ષેપમાં સ્પષ્ટ કરવાની મોટે ભાગે જરૂર પડે છે. દા.ત. 'Information Retrieval' ના સર્ચ માટે કાં તો આ પ્રમાણે વ્યક્ત કરી શકાય.

information AND Retrieval

અથવા તો

Retrieval AND Information

એ ડોક્યુમેન્ટ કે જે 'Information' અને રીટ્રાઈવલ બંને ધરાવે છે તેને રીટ્રાઈવ કરવાને બદલે ડોક્યુમેન્ટના જરૂરી સેટ

---

(1) વિવિધ સર્ચ એન્જિન માટેની સૂચનાઓ નોંધવી જરૂરી છે. ગૃપ સર્ચ ટર્મ માટે પેરેન્થિસિસનો ઉપયોગ કરવામાં આવે છે એની પણ નોંધ લેવી જોઈએ કે બુલિયન ઓપરેટર્સ જુદી સર્ચ લેંગ્વેજ ઉપર જુદા નોટેશનનો ઉપયોગ કરીને જુદી રીતે પ્રતિનિધિત્વ કરે છે. કમનસીબે

કોઈ પણ ચોક્કસ માપદંડ સુસંગત રીતે અનુસરવામાં આવતું નથી. ઘણાં સર્ચ એન્જિન્સ પણ કેસ - સેન્સીટીવ હોય છે. એકલ ફોર્મનો ઉપયોગ બહુધા ફોર્મમાં રીટ્રાઈવ ન થઈ શકે. સર્ચ એક્સપ્રેશન ધરાવે છે, તે અંકોના નંબરની પણ મર્યાદા હોય છે. કે જ્યાં "Information" શબ્દ "Retrieval" શબ્દની તરત જ આગળ આવે છે.

(b) ટર્મના મોર્ફાલોજીકલ વેરીએન્ટને આવરી લેતા બુલિયન ઓપરેટર્સનો ઉપયોગ કરવો કંટાળાજનક છે. (દા.ત. Classify of Classification or Classifying or Classing)

(c) બુલિયન ઓપરેટર AND બે સર્ચ ટર્મ વચ્ચેના ઈચ્છિત સંબંધ વ્યક્ત કરવા માટે અપૂરતો છે.

(d) બુલિયનનું સર્ચીંગ મોડેલ આઈટમને નિર્દિષ્ટ કરે છે કે તે ક્વેરીના ડોક્યુમેન્ટ / રેકર્ડમાં પ્રસ્તુત છે કે નથી. એ બધા જ ડોક્યુમેન્ટને સમાન અગત્યતા આપે છે. આ પ્રમાણે રીટ્રાઈવ્ડ કરી મીકેનીઝમને કોઈ પણ પ્રકારની ક્વેરીને લગતા કેટલીક સમસ્યાઓ પર આધારિત આઉટપુટને રેન્ક આપવા માટે પુરૂ પાડતું નથી.

આ મર્યાદાને દૂર કરવા માટે સૂચવાયેલી એક રીત છે (Best match Serching) કે જે રીટ્રાઈવ્ડ રેકોર્ડને વજન આરોપણ કરવા માટે સંભાવનાની થિયરીનો ઉપયોગ કરે છે. આનો સમાવેશ કેટલાક ક્વોન્ટીટેટીવ મેઝર ઓફ સિમિલારિટા ડેટાબેઝમાં ક્વેરી અને ડોક્યુમેન્ટ / રેકોર્ડ અને રેકર્ડ આઉટપુટને ઉત્પન્ન કરે છે. વિદ્યાર્થીઓએ નોંધ લેવી જોઈએ કે જે www. સર્ચીંગ કરે છે તેઓ 'hits' ના રેંક્ડ આઉટપુટ પેદા કરે છે એ ડેટાબેઝનો વિષય છે કે તે આવા રેન્કીંગ ચર્ચાને પાત્ર વિષયમાં કેટલા વિશ્વાસપાત્ર છે. આ વિશે આપણે વધારે માહિતી આગળના પ્રકરણમાં જોઈશું.

◆ **તમારી પ્રગતિ ચકાસો**

3. Venn (વેન) ડાયાગ્રામનો ઉપયોગ કરી બુલિયન ઓપરેટર્સને સમજાવો

નોંધ (i) તમારા જવાબ નીચે આપેલી જગ્યામાં લખો

(ii) તમારા જવાબો આ પ્રકરણના અંતે આપેલા ઉત્તરો સાથે સરખાવો

....

....

....

....

....

....

4.7 પ્રગતિશીલ ફિચર્સ (ADVANCED FEATURES)

આપણે અત્યાર સુધી સર્ચની કેટલીક કી-વર્ડ્ઝ અને વાક્યોના સંયોગ સાથે જોડાયેલી શક્તિને જોઈ. ઘણાં સર્ચ એન્જિન ઘણી પ્રગતિશીલ શક્તિ ધરાવે છે. એ યાદ રાખવું અગત્યનું છે કે, આ બધી પ્રગતિ એ આ સર્ચ એન્જિન થકી ઉપલબ્ધ નથી થતી અને કેટલીક તો એડવાન્સ ફીચર આપતા નથી. આ પછીના પેરેગ્રાફમાં આપણે આવા પ્રગતિશીલ ફિચર્સને સંક્ષેપમાં તપાસીશું.

4.8 કીવર્ડ ગ્રુપીંગ (KEYWORD GROUPING)

જટીલ સંયુક્ત ક્વેરીમાં કે જેને અનેક કીવર્ડ્ઝ અને / અથવા વાક્યો હોય છે અને તેમને કી વર્ડ્ઝ / વાક્યોને અન્ય બીજા કી-વર્ડ / વાક્યો સાથે ગૃપ કરવા અથવા લીંક કરવા જરૂરી હોઈ શકે.

ઉદાહરણ નીચેના સર્ચને તપાસો

UK AND Football OR USA AND Soccer

રીટ્રાઈવીંગ ડોક્યુમેન્ટને કે જે Football in U.S.A. પર છે, તે ખરેખર અહીં અપ્રસ્તુત છે. તેને બાકાત રાખવા માટે મેથેમેટિકલ ઓપરેશન્સની જેમ સર્ચ ટર્મના ગૃપનું હોવું અગત્યનું છે. કી-વર્ડઝને ગૃપ કરવા માટે પેરેનથસીસનો ઉપયોગ કરવામાં આવે છે.

(UK AND Football) OR (USA AND Soccer)

4.9 ટ્રંકેશન (TRUNCATION)

ટ્રંકેશન એ ઘણા સર્ચ એન્જિન પૈકી બહુ જ શક્તિશાળી ફિચર્સ છે. અગાઉ 'OR'નો ઉપયોગ કરી લીંકીંગ કંટાળાજનક છે તે ઉદાહરણને ધ્યાનમાં રાખીને વધારે માત્રામાં કી વર્ડઝ કે જેમાંની દરેક માર્કોલોજી વેરિઅન્ટ છે તેનો ઉપયોગ કરવામાં આવ્યો છે. તેથી ક્વેરીને ઇનપુટ કરવાનું નીચે પ્રમાણે સરળ થશે.

Classic*

જો કે વાઈલ્ડ કાર્ડનો ઉપયોગ કરતી વખતે કાળજી રાખવાની જરૂર છે.

Classic, Classics, Classy Classis તેથી ક્લાસીક, ક્લાસીક્સ, ક્લાસી ક્લાસીક વગેરે હોય તેવી ક્વેરી પણ ડોક્યુમેન્ટને રીટ્રાઈવ કરી શકશે કે જે ખરેખર અપ્રસ્તુત છે. ઘણાં સર્ચ એન્જિન (ખાસ કરીને ઈન્ટરનેટ માટે વ્યાપક પ્રમાણમાં ઉપયોગમાં આવતા) વાઈલ્ડ કાર્ડથી શરુ થતી ક્વેરીને ટેકો આપતા નથી. આમ છતાં સર્ચ લેંગ્વેજીસ ઘણી છે (દા.ત. કોમર્શીયલમાં ઉપયોગમાં આવતી ડેટાબેઝીસ) તે કોઈ પણ બિંદુએ કેરેક્ટરના સ્ટ્રીંગને અંતે શરુઆતમાં બે સ્ટ્રીંગની વચ્ચે, કેરેક્ટરના સ્ટ્રીંગના અંતે, શરુઆતમાં અને કેરેક્ટરના સ્ટ્રીંગના અંતે, ટ્રંકેશનના ઉપયોગને સમર્થન આપે છે.

4.10 પ્રોક્સીમીટી ઓપરેટર્સ (Proximity Operators)

આ ઓપરેટરો બુલિયન ઓપરેટર AND સર્ચરને બે સર્ચ ટર્મ વચ્ચે મહત્તમ છૂટ આપી શકાય તેવું અંતર નક્કી કરવા માટે છૂટ આપે છે. પ્રોક્ષીમીટી ઓપરેટર (સંનિધિ ઓપરેટર તરીકે પણ ઓળખાય છે.) વિચારાધિન પ્રસ્તુત આઈટમ માટે બે સર્ચ ટર્મ એક પછી એક આપવી જોઈએ કે મહત્તમ શબ્દો અને સર્ચરને બે સર્ચ ટર્મ વચ્ચે નક્કી કરવાની છૂટ આપે છે. રીટ્રાઈવલ સિસ્ટમની ક્વેરીમાં વપરાતા પ્રોક્ષીમીટી ઓપરેટર્સના ઉદાહરણ નીચે પ્રમાણે બતાવ્યા છે.

ઉદાહરણ

BRSમાં નીચેની ક્વેરી : (Information with Retreival)

બધાં જ ડોક્યુમેન્ટ રીટ્રાઈવ કરશે કે જેમાં એ જ વાક્યમાં બે કી વર્ડઝ હોય.

IN WINSIS THE QUERY

INFORMATION. RETREIVAL

એ બધાં ડોક્યુમેન્ટ રીટ્રાઈવ કરશે જેમાં બે કી વર્ડઝ આવતા હોય કે જેમાંથી એકબીજાનો શબ્દ એ જ વાક્યમાં આવતો હોય.

4.11 લીમીટીંગ સર્ચ (Limiting Search)

ફિલ્ડ સ્પેસિફીક સર્ચ : ફિલ્ડ સ્પેસિફીક સર્ચીસ એ મોટા ભાગની સિસ્ટમમાં સમાન ફિચર્સ છે જ્યારે સર્ચરને ચોક્કસ ફિલ્ડમાંથી કી વર્ડને સર્ચની મર્યાદામાં બાંધવાની જરૂર હોય છે ત્યારે ખાસ કરીને આ મદદરૂપ થાય છે. દા.ત. ટાઈટલમાં અથવા ડિસ્ક્રીપ્ટર ફિલ્ડમાં ઘણાં સર્ચ એન્જિન, વધારાના પેરામીટર્સને ઉમેરવાની છૂટ આપે છે. પણ તે કી વર્ડઝ નથી પણ સર્ચર જેને સર્ચ કરવા માંગતું હોય

તેને વધારાની માહિતી પૂરી પાડે છે. કેટલાક ફિલ્ડ કે જેના દ્વારા સર્ચીસ મર્યાદિત કરવામાં આવ છે તેને સર્ચ એન્જિન દ્વારા સમર્થન કરવામાં આવે છે તેમાં સમાવેશ થાય છે.

- Publication Date/ Range
- Language
- Country of Publication
- Type of File (e.g. word Document, PDF File, Mp3 etc.)
- Sorting Criteria
- **સ્ટ્રીંગ સર્ચિંગ :** આ સગવડ વપરાશકર્તાને કેરેક્ટરના સ્ટ્રીંગને ટેકસ્ટના ભાગને પ્રત્યારોપણ કરવા માટે મંજૂરી આપે છે. આ સગવડ ખાસ કરીને કેરેક્ટર સ્ટ્રીંગને કે જે ઇન્ડેક્ષ થનાર ન હોઈ શકે અથવા ઈન્ડેક્ષ થયેલા ન હોય તેને સર્ચ કરવા માટે ઉપયોગી છે. દા.ત. લાયબ્રેરી કેટલોગમાં સામાન્ય રીતે પ્રકાશકનું નામ ઇન્ડેક્ષ થયેલું હોતું નથી. ચોક્કસ પ્રકાશક દ્વારા પ્રકાશિત થયેલ લાઇબ્રેરીમાં આઇટમને રીટ્રાઇવ કરવાની જરૂર પડે તો સ્ટ્રીંગ સર્ચિંગનો ઉપયોગ કરી શકાય. સર્ચ એન્જિન કેરેક્ટરના સ્ટ્રીંગની હાજરીને (પ્રકાશકનું નામ નિર્દેશ કરીને) યોગ્ય ફિલ્ડમાં ડેટાબેઝના દરેક રેકોર્ડમાં શોધશે દા.ત. ઇમ્પ્રીંટ ફિલ્ડ (પ્રકાશક, મુદ્રક વગેરેના નામ સરનામાની વિગતો)

Webના સર્ચિંગમાં મુખ્ય પ્રશ્ન ખાસ કરીને Web ના સર્ચિંગનો છે. આ પ્રશ્નને નજરમાં રાખીને કેટલાક સર્ચ એન્જિનને ઉતરતી માત્રામાં ક્વેરીઝને અનુરૂપ પરિણામને સોર્ટ કરવા માટે સગવડ પૂરી પાડે છે, કે જે આઇટમને વધારે પ્રસ્તુત અને અનુરૂપ સાબિત થઈ હોય તેને આ આઉટપુટના લિસ્ટમાં સૌથી ઉચ્ચ સ્થાને મૂકવામાં આવે છે. સોર્ટિંગ એ ખૂબ શક્તિશાળી રસ્તો છે કે જે ખરેખર કોઈ અનુરૂપને બાકાત રાખતું નથી. આમ છતાં હાલમાં થોડાક સોર્ટિંગ માનદંડ જો કોઈ હોય તો તેને થોડાક Web સર્ચ એન્જિન દ્વારા સમર્થન આપવામાં આવે છે. ઘણાં પરંપરાગત IRS માં વેઇટેડ ટર્મ સર્ચિંગના ફોર્મમાં કેટલીક સર્ચ લેંગ્વેજીસમાં સગવડ પ્રાપ્ય છે ખાસ કરીને આ સગવડ ‘Weight’ એ દરેક રીટ્રાઇડ આઇટમને તેની ગણત્રીની માત્રામાં ક્વેરીને અનુરૂપ અને આઉટપુટ વિષયને લગતા ઉતરતા ક્રમમાં સાંકળે છે. વધારે પ્રસ્તુત ડોક્યુમેન્ટ એ આઉટપુટ લિસ્ટના ઉચ્ચ સ્થાને છે. આમ છેવટના વપરાશકર્તા સમયની બચત કરે છે. કારણ કે તેને પરિણામનું લાંબુ લિસ્ટ, લિસ્ટના અંતમાં આપેલી વધારે પ્રસ્તુત આઇટમને તપાસવાની જરૂર નહીં પડે.

- **તમારી પ્રગતિ ચકાસો**

4 નીચે આપેલા વ્યાપક પ્રમાણમાં ઉપયોગમાં આવતા IR પેકેજમાં પૂરા પાડેલા એડવાન્ડ્સ (આધુનિક) સર્ચફિચર્સનું લીસ્ટ આપો

(i) WINSIS

(ii) WINSPIRS

(5) નીચેની વેબ એન્જિનમાં પ્રાપ્ય એડવાન્સ્ડ સર્ચફિચરનું લીસ્ટ બનાવો અને તેની સરખામણી કરો.

(i) Atlavista

(ii) Alltheweb

(iii) Hotbot

(iv) Google

નોંધ (i) તમારા જવાબો નીચે આપેલી જગ્યામાં લખો.

(ii) પ્રકરણના અંતે આપેલા ઉત્તરો સાથે સરખાવો

....

....

....

....

....

....

4.12 ટ્રેન્ડઝ ઇન ઇન્ફર્મેશન રીટ્રાઇવલ (Trends in Information Retrieval)

કોમ્પ્યુટર અને કોમ્યુનિકેશન ટેકનોલોજીના વિકાસે માહિતીના સ્ટોરેજ અને ટ્રાન્સફરના પ્રશ્નો સરળ કર્યાં છે. તેણે સર્ચિંગ અને માહિતીના રીટ્રાઇવલનું કામ જટીલ બનાવ્યું છે. વિસ્તીર્ણ ઓનલાઇન ડેટાબેઝના આગમન સાથે છેલ્લા થોડા દાયકાઓમાં વધતી જટિલતાને લીધે ઉભા થતા કેટલાક પ્રશ્નોના નિરાકરણ માટે IR રીસર્ચ માટેના પ્રયત્નો થઈ રહ્યા છે જે વિસ્તારમાં સંશોધન આગળ ચાલી રહ્યું છે તે બે કે ત્રણ મુખ્ય વિસ્તાર છે. આ બધા વિસ્તારનું નિરીક્ષણ (Over view) સંક્ષેપમાં આપવા કઠીન છે ત્યારે મુખ્ય વલણનો સામાન્ય વિચાર જાણવો અગત્યનો છે.

◆ **Intelligent IR ઇન્ટેલિજન્ટ આઇ. આર :** એવાં પ્રયત્નો થઈ રહ્યા છે કે કોમ્પ્યુટર માનવની જેમ વિચારી શકે. કેટલાક એક્સપોર્ટ સિસ્ટમના પ્રયોગ (eg. MYCLN) અને આર્ટિફિશીઅલ ઇન્ટેલીજંટના કેટલાક ફિચરનો IR સીસ્ટમમાં સમાવેશ એ નોંધવું અહીં જરૂરી છે. કેટલાક સર્ચ એન્જિન્સ ખાસ કરીને Webને સર્ચ કરવા માટે ક્વેરીઝને નેચરલ લેંગ્વેજ ક્વેરીના ફોર્મમાં મેળવવા માટેની સગવડ માટે પ્રયત્ન કરવામાં આવે છે. દા.ત. 'ચાઇનાની વસ્તી કેટલી છે ?' સર્ચ એન્જિન તે પછી નેચરલ લેંગ્વેજ ક્વેરીને અલગોરિધમમાં બદલશે, આ સાધનના ઘણાં પ્રશ્નો છે. અને તે કેટલાક સરળ / અથવા સમાન ક્વેરીઝ માટે કાર્ય કરે છે. પણ ઓછા સામાન્ય ટાઇપનીક્વેરી માટે કામ કરતા નથી (Belkin)ની Ask (Amomalous State of Knowledge) કલ્પિતાર્થથી (Hypothesis) શરૂ કરીને IR સિસ્ટમ માટે વધારે સારા End User મોડેલ્સ અને ઇન્ટરફેસ માટે વિકાસ કરવા માટે પ્રયત્નો કરવામાં આવે છે. યુઝર ઇન્ટરફેસીસે ઘણીં પ્રગતિ કરી છે. ખાસ કરીને ઘટકને સંબંધિત મેનુ, ફોર્મ્સ, ગ્રાફિક્સ, હાઇપર ટેક્સ્ટ વગેરે IRમાં નેચરલ લેંગ્વેજ પ્રોસેસીંગ (NLP) નેચરલ લેંગ્વેજ ટેક્સ્ટ (ડોક્યુમેન્ટના ટેક્સ્ટ તથા એન્ડ યુઝર્સ ક્વેરી) જે કોમ્પ્યુટરમાં સ્ટોર કરેલી છે તેના ફોર્મમાં વધારે યોગ્ય પ્રક્રિયા માટે બદલવા માટે મેનીપ્યુલેટ કરી શકાય દા.ત.

- ડિઝાઇનીંગ NL ઇન્ટરફેસીસ કે જે નેચરલ લેંગ્વેજ ક્વેરીના ફોર્મમાં ક્વેરીઝને સ્વીકારે છે.
- સ્ટોરેજ સાઇડમાં (NL) ટેક્સ્ટ પ્રોસેસિંગ કે જે ટેક્સ્યુઅલ ડેટાના મોટા ભાગને ઓટોમેટિકલી ડીરાઇવ નોલેજ સ્ટ્રક્ચરને ટેકો આપે છે. (દા.ત. ઓટોમેટિક વર્ગીકરણ) તેનો ઉપયોગ ટેક્સ્ટમાંથી માહિતી મેળવવાથી અને રીટ્રાઇવ કરવા માટે પ્રગતિમાં છે. વિકાસનો અર્થ સૂચન કરે છે કે IR રીસર્ચ એ ઇન્ટેલિજન્ટ ઇન્ફોર્મેશન પ્રોસેસિંગ ટેકનોલોજીને જેવી કે પરંપરાગત બુલિયન IR મોડેલના ઉપયોગની ઘણી મર્યાદાને સુધારવા માટેનો પ્રયાસ કરે છે. હાલના વર્ષોમાં IR રીસર્ચ એ ઘણું આનુષાંગિક વિસ્તાર સાથે મજબૂત લીંક સ્થાપિત કરી રહી છે. દા.ત. NLP આર્ટિફિશીઅલ ઇન્ટેલીજન્સ, HCL કોગ્નિટીવ સાયન્સીસ વગેરે. (જ્ઞાનને લગતું વિજ્ઞાન) આમ છતાં એ યાદ રાખવું જરૂરી છે કે, આમાંના ઘણાં સંશોધનો પ્રયોગાત્મક તબક્કામાં છે અને મોટા સ્ટ્રક્ચર્ડ કોમર્શિયલ ડેટાબેઝ આ ટેકનોલોજીને અપનાવે તેને થોડો સમય લાગશે.

4.13 સારાંશ (SUMMARY)

આ પ્રકરણમાં તમને ડેટાબેઝ સર્ચીંગમાં અને વિવિધ સાધનોમાં જે સર્ચને વધારવા માટે પ્રાપ્ય હોય છે તે વિશેની જાણકારી આપવામાં આવી છે તમે એ પણ શીખ્યા હશો કે વિવિધ સર્ચ સાધનોની રીકોલ અને સર્ચ આઉટપુટના પ્રીસીઝન ઉપર શું અસર થાય છે. આવા સર્ચ સાધનોનો વધારે અભ્યાસ કરવા માટે વધારે યોગ્ય રસ્તો એ છે કે સર્ચ લેંગ્વેજને અને સર્ચ સાધનોનો વિવિધ જોડાણમાં ઉપયોગ કરવો. રજૂ કરવામાં આવેલી કેટલીક એક્સરસાઇઝ આની તરફ નિશાન તાકવામાં આવ છે અને આશા રાખવામાં આવે છે કે વિદ્યાર્થીઓ કેટલાક સર્ચ એન્જિન્સ અને સચ સાધનોનો ઉપયોગ કરીને પ્રયત્ન કરે અને તેમના ઉપયોગ માટે પ્રત્યક્ષ અનુભવ પ્રાપ્ત કરે.

4.14 તમારી પ્રગતિ ચકાસોના ઉત્તરો (ANSWERS TO SELF CHEK EXCERCISES)

1. ટર્મ ઇન્ફોર્મેશન રીટ્રાઇવલી Calvin Mooers એ ઉપજાવ્યો હતો. તે ડેટાબેઝમાંથી ક્વેરીમાં પ્રસ્તુત રીટાઇવિંગ ડોક્યુમેન્ટને મેળવવા માટેની પ્રક્રિયા છે. વ્યાપક રીતે જોઈએ તો IRમાં સંકળાયેલી વિવિધ પ્રક્રિયા, જેમાં ડોક્યુમેન્ટના ડેટાબેઝની પ્રક્રિયા અને સર્ચ સાથે જે સંકળાયેલા છે, તેનો સમાવેશ થાય છે.

2. ઈન્ફોર્મેશન રીટ્રાઇવલ સિસ્ટમને (ટેક્સ્ટ રીટ્રાઇવલ સીસ્ટમ તરીકે પણ ઓળખાય છે.) ડોક્યુમેન્ટની ટેક્સ્ટને મેનેજ કરવા માટે જ્યારે IRS નું રેકોર્ડ સ્ટ્રક્ચર ફિલ્ડમાં સ્ટ્રક્ચરીંગ ડેટાને સહાય કરે છે, ત્યારે DBMS થી તે જૂદું પડે છે, કે જે હાઇલી સ્ટ્રક્ચર્ડ ડેટાને મેનેજ કરવા માટે રચવામાં આવ્યું છે. દા.ત. રીલેશનલ DBMS જેવા કે INGRES હાઇલી સ્ટ્રક્ચર્ડ અને ડેટાના ઇન્ટરરીલેટેડ બળને મેનીપ્યુલેટ કરે. આવા DBMS ચોક્કસ લાઇબ્રેરી ઓપરેશન્સ જેવા કે સરક્યુલેશન માટે યોગ્ય હશે, પરંતુ તે સ્ટોરીંગ ટેક્સ્ટ / ઓબસ્ટ્રેક્ટ ઓફ ટેક્સ્ટ વગેરેમાં મર્યાદા બાંધે છે કે જે ટેકસ્ટ રીટ્રાઇવલ સીસ્ટમમાં જરૂરી છે. અપેક્ષિત આઉટપુટની ટર્મમાં પણ એક મહત્ત્વનો તફાવત છે. ડેટા રીટ્રાઇવલમાં આપણે રેકોર્ડને શોધીએ છીએ કે જે ક્વેરી માટે ચોક્કસ અનુરૂપ હોય છે. જ્યારે ઇન્ફોર્મેશન રીટ્રાઇવલમાં આપણે રેકોર્ડને રીટ્રાઇવ કરવા માટે ક્વેરીની ડીગ્રી ઓફ સિમિલારિટી ઉપર આધાર રાખીએ છીએ.

3. કૃપયા આ બાબત ઇન્ફોર્મેશન રીટ્રાઇવલની કોઈ એક સારી ટેક્સ્ટ બુકને રીફર કરો.

4. સર્ચ સાધનો (Tools) જેવા કે WINSPIRS અને WINSIS સર્ચ લેંગ્વેજ, સર્ચ ડિવાઇસીસની વિશાળ શ્રેણી જેવી કે, બુલિયન ઓપરેટર્સ, પ્રોક્ષિમીટી ઓપરેટર્સ, ટ્રંકેશન, ફિલ્ડ સ્પેસિફીક સર્ચિંગ, ફ્રી ટેકસ્ટ સર્ચિંગ વગેરે, આમ છતાં તેની મર્યાદા છે. આની સંપૂર્ણ માહિતી માટે પ્રસ્તુત મેન્યુઅલનેરીફર કરો.

5. બુલીઅન ઓપરેટર્સને (Web) સર્ચ એન્જિન દ્વારા સમર્થન મળે છે. તેમાંના કેટલાક પ્રોક્ષિમીટી ઓપરેટર્સ, ફ્રેઝસર્ચિંગ અને તે ઉપરાંત વેબ સર્ચિંગના અનુસંધાને સર્ચ ફીચરને પણ પુરું પાડે છે. (દા.ત. લીંક સર્ચિંગ) આમ છતાં ટ્રંકેશન એ વેબ સર્ચ એન્જિન દ્વારા આધારિત ફિચર નથી. Greg Notess એ વ્યાપક પ્રમાણમાં સારી રીતે ઉપયોગમાં આવતા સર્ચ એન્જિનનું પૃથક્કરણ કર્યું છે.

(Visit Searchengineshowdown.com you can also get to these from the home page of Gerg Notess www.notess.com)

4.15 ચાવીરૂપ શબ્દો (KEYWORDS)

Boolean Operators : A system of three symbolic logical operators to combine terms to constitute a search statement, representing * *sum** (+) **(OR),** *'logical product** (x) (AND) and *'logical difference'* (-) (NOT)

Data : Raw numbers or letters stored in a computer which represent about entities

[2] Students are advised to note that the term *'Relevance'* in information retrieval is far from being clearly defined. Tefko Saracevic has discussed the many definitions of *'relevance'*.

Database : An organised collection of data that exists for the purpose of providing information.

Descriptor : A word / term taken from a controlled vocabulary - usually a thesaurus - and assigned to a text to represent, partially or fully, the ***'aboutness'*** of the text *(see also Keyword)*

Field Specific search : *A* search facility that allows the searcher to specify the field (e.g. Title field) in which a search term should be present

Information Retrieval : The process of retrieving information about documents in a database relevant to a query

Keyword : Technically, a significant word / term taken from the text of a document *(see, also Descriptor)*

Proximity Operator : A set of operators to combine search terms that permit the searcher to specify the context in which a search term should occur. Proximity operators allow the searcher to specify whether two search terms should occur adjacent to one another or the maximum number of other words that could be permitted between the search terms

Search Engine : Large databases of web resources along with a search language of their own to support a wide range of search features to search the databases.

Search Language : An artificial language used to construct search statements that can be input to an IRS;

Truncation : A search device that supports searching of records containing any one or more of a number of different terms by merely specifying a string of characters that is common to the different search terms. The common string of characters could be in the beginning (root) or at the end or even in the middle of a word; widely used for words having morphological variations

4.16 સંદર્ભ અને વિશેષ વાચન (REFERENCE AND FURTHER READING)

Information retrieval is considered the essence of Information Science. There is therefore no dearth of good reading material and standard textbooks. **G.G. Chowdhury's** book referred to in this Unit is a comprehensive textbook covering adequately all aspects of database searching including web searching. Reference has also been made to the website of **Greg Notess** that provides useful material especially on web searching and search engines. **F. W. Lancaster** has written standard texts on information retrieval. Besides these, there are other useful materials that students will benefit from:

Ashford, J. and Willet, P. (1988). *Text Retrieval and Document Databases.* Bromley: Chartwell-Bratt.

Bates, Marcia J. (1981). Search Techniques. *Annual Review of Information Science and Technology.* 16, "139-169.

Ingwersen, P. (1996). Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation,* 52(1), 3-50.

Meadow, C.T. and Cochrane. P. A. (1981) *Basics of Online Searching.* New York: Wiley.