એકમ–18 માહિતી પુનઃ પ્રાપ્તિ મોડેલ અને તેઓનું અમલીકરણ
INFORMATION RETRIEVAL MODELS AND THEIR APPLICATIONS

: રૂપરેખા :

18.0 ઉદ્દેશો
18.1 પ્રસ્તાવના
18.2 માહિતી પુનઃપ્રાપ્તિ
18.2.1 ડેટા સંપુટ
18.2.2 માહિતી આધાર
18.2.3 માળખાગત શોધ ભાષા
18.3 માહિતી પુનઃપ્રાપ્તિની તક્નિકો
18.4 મોડેલ આધારિત નિવેશ/નિર્ગમન
18.4.1 ડેટા પુનઃ પ્રાપ્તિના મોડેલ
18.4.2 માહિતી પુનઃ પ્રાપ્તિના મોડેલ
18.4.3 જ્ઞાન પુનઃ પ્રાપ્તિના મોડેલ
18.5 મોડેલ આધારિત સિદ્ધાંતો અને સાધનો
18.5.1 બુલીયન પુનઃ પ્રાપ્તિ મોડેલ
18.5.2 અસ્પષ્ટ તાર્કિક નમુનાઓ
18.5.3 ગણ સિદ્ધાંત આધારિત મોડેલ
18.5.4 સદિશ અવકાશ મોડેલ
18.5.5 સંભાવનાત્મક પુનઃપ્રાપ્તિ મોડેલ
18.5.6 ભાષાશાસ્ત્રીય મોડેલ
18.5.7 ગાણિતીક મોડેલ
18.5.8 મનોવૈજ્ઞાનિક મોડેલ
18.5.9 આર્થિક મોડેલ
18.5.10 ઉચ્ચકક્ષીય પાઠ જોડાણ નમૂનો
18.6 સારાંશ
18.7 તમારી પ્રગતિ ચકાસોના જવાબો
18.8 ચાવીરૂપ શબ્દો
18.9 સંદર્ભો અને વિશેષ વાચન

18.0 ઉદ્દેશો (OBJECTIVES)

માહિતી પુનઃપ્રાપ્તિની યોગ્ય તક્નિકોની શોધ માટેના પ્રયત્નોમાં વિવિધ મોડેલ વિકસાવવામાં આવેલ છે. આ એકમમાં માહિતી પુનઃપ્રાપ્તિની તક્નિકો અને સમાંતરે વિકસાવવામાં આવેલ મોડેલ્સની ચર્ચા કરવામાં આવેલ છે. આ એકમના અધ્યયન બાદ તમે આ બાબતોથી સક્ષમ બનશો.

- આ માહિતી પુનઃ પ્રાપ્તિમાં સમાવેશ થતી પાયાગત બાબતો અને પરિબળો જાણી શકશો.
- માહિતી પુનઃ પ્રાપ્તિની પ્રણાલિગતથી આધુનિક તરફ થયેલ પરિવર્તન જાણી શકશો.
- માહિતી જરૂરિયાત ડેટાસંપુટો, જ્ઞાન આધારો, માહિતી પદ્ધતિઓ અને ગ્રંથાલયોમાંથી માહિતી પુનઃ પ્રાપ્તિની મેળવણીની પ્રક્રિયા સમજી શકશો.
- માહિતી પુનઃ પ્રાપ્તિ મોડેલના વિકાસથી માહિતીગાર થઈ શકશો.
- માહિતી પુનઃ પ્રાપ્તિના સંશોધન ક્ષેત્રોથી તમારી જાતનો પરિચય પ્રાપ્ત કરી શકશો.

18.1 પ્રસ્તાવના (INTRODUTION)

માહિતી પુનઃ પ્રાપ્તિનું વલણ વ્યક્તિઓના સમૂહોની જ્ઞાનની શોધની વર્તણૂંકનું નિદર્શન કરે છે. આ સંદર્ભમાં શોધક જ્ઞાનના વિશાળ સંગ્રહમાંથી કેટલીક માહિતી શોધે છે. મનની આ સ્થિતિનું પૃથક્કરણ અને નિદાન ગ્રંથાલયો, માહિતી પુનઃ પ્રાપ્તિ પધ્ધતિઓ, ડેટા સંપુટો, જ્ઞાનના આધાર અને તેવા જ પ્રકારના વાતાવરણમાં જ્ઞાનના સંગઠન માટે માર્ગદર્શન પૂરું પાડે છે. આ માર્ગદર્શિકાઓનો હેતુ શોધકના અભિગમ અને ડેટા સંપુટમાં જ્ઞાનના સંગઠન વચ્ચે સુમેળતા પૂરી પાડવાના સમયે મદદરૂપ થવાનો હોય છે. માનવનું શિક્ષણ, સમસ્યાનું નિવારણ કરવાનું અને નિર્ણય કરવાની પરિસ્થિતિનો સમાવેશ કરતું સાંપ્રત વાતાવરણ દરેક ઘટના પ્રસંગે જ્ઞાનના માળખામાં લચીલાપણાની માગણી કરે છે. કમ્પ્યુટર અને પ્રત્યાયન ટેકનોલોજીમાં થયેલ વિકાસે વિશાળ જથ્થાવાળી માહિતીને સઘન સ્વરૂપે સંગ્રહ કરવાનું શક્ય બનાવેલ છે. વિકસાવવામાં આવેલ વૈવિધ્યતાયુક્ત સોફ્ટવેર આ સંગ્રહમાંથી ઝડપી પુનઃ પ્રાપ્તિ કરવા માટેનો અવકાશ પૂરો પાડેલ છે. જ્યારે માહિતી પુનઃ પ્રાપ્તિની ઝડપ મૂલ્યવાન છે ત્યારે જો તે માહિતી શોધકને તૈયાર ઉપલબ્ધ કરી શકાતી હોય તો તે વધુ સમૃધ્ધ બનાવી શકાય છે. આ સંદર્ભ માહિતી પુનઃપ્રાપ્તિના મોડેલ, ઉપભોક્તા મૈત્રીપૂર્ણ અભિગમમાં બૌદ્ધિક માહિતી પુનઃપ્રાપ્તિના મોડેલ તૈયાર કરવાની માત્રા દર્શાવે છે.

આ પ્રકારના વિકાસે બૌદ્ધિક વિજ્ઞાન ક્ષેત્રોનો પ્રારંભ કરાવ્યો છે જે આંતર વિદ્યાશાખાવાળું છે જેમાં મનોવિજ્ઞાન વર્તણૂકીય અભ્યાસો, કમ્પ્યુટર વિજ્ઞાન, કૃત્રિમ બુદ્ધિ અને માહિતી વિજ્ઞાનના ક્ષેત્રોમાંથી વિચારોને નિવેશ માટે ખેંચી લાવવામાં આવે છે.

18.2 માહિતી પુનઃપ્રાપ્તિ (INFORMATION RETRIEVAL)

માહિતી પુનઃ પ્રાપ્તિ એ સંગ્રહમાંથી માહિતી પસંદ કરવાની પ્રક્રિયા છે. તેનો અર્થસૂચિત કરે છે કે માહિતી માટેની શોધ પ્રલેખ, મેટા ડેટા જે પ્રલેખો વર્ણવે છે અથવા અંતરિયાળ ડેટા સંપુટોમાંથી હોઈ શકે છે. ડેટા સંપુટો ફક્ત ડેટા સંપુટો અથવા હાઈપર ટેકસ્ટ ધરાવતા નેટવર્કીંગ ડેટા સંપુટ જેવા કે ઈન્ટરનેટ અથવા ઈનપુટ હોઈ શકે છે. પ્રાથમિક રીતે તે એન્ટ્રી વ્યક્તિઓને મદદ કરે છે કે જેઓને તેમની સમસ્યા નિવારણ, નિર્ણય પ્રક્રિયા, ઉત્પાદન સેવા જેવી પ્રવૃત્તિઓમાં કેટલીક માહિતી મેળવવા શોધ જરૂરી હોઈ શકે, વ્યાપક રીતે કહીએ તો ચાર પ્રકારની માહિતી પુનઃ પ્રાપ્તિ અસ્તિત્વ ધરાવે છે. પ્રલેખ પુનઃપ્રાપ્તિ જેમાં સાદું માળખું ધરાવતી ફાઈલની દરેક નોંધના થોડીક સંખ્યામાં સુવ્યાખ્યાયિત કરેલ લક્ષણો અને મર્યાદિત રીતે પૂર્વ નિર્ધારિત શોધ પ્રશ્ન પ્રકારોના સમૂહનું ડેટા સંપુટમાંથી માહિતી પુનઃપ્રાપ્ત થાય તે રીતે પ્રક્રિયા કરવામાં આવે છે. સંદર્ભ પુનઃપ્રાપ્તિ જેમાં નોંધો-પુસ્તકો, પ્રલેખો અને ગ્રંથાલયની અન્ય સામગ્રીનું નિરૂપણ કરે છે અને સંખ્યાબંધ જુદા જુદા લક્ષણોવાળી બાબતોની અમર્યાદિત અસરકારક ઓળખ પ્રાપ્ત હોય, તેવા પ્રસંગોમાં શોધ પ્રશ્ન ઘણીવાર પ્રલેખની વિષયવસ્તુનો સંદર્ભ આપે છે. ઘણા સામાન્ય પ્રસંગોમાં પુનઃપ્રાપ્તિની રચના કોઈપણ પ્રકારના શોધપ્રશ્નને હસ્તગત કરવા હોઈ શકે છે અને પધ્ધતિ આવા શોધ પ્રશ્નો પ્રત્યક્ષ જવાબ આપી પૂર્ણ કરી શકે છે. હકીકત પુનઃ પ્રાપ્તિમાં વિશાળ સંખ્યામાં વૈવિધ્ય ધરાવતા જુદા જુદા માહિતી ઓળખોની જરૂરિયાત હોઈ શકે છે. પાઠ પુનઃપ્રાપ્તિમાં સંદર્ભ,થવા ડેટા અથવા સમગ્ર પ્રલેખોની સમૂહ પુનઃપ્રાપ્તિ કરવાને બદલે કોઈ એક ખાસ વિષય એકમ ઉપરનો પાઠ પુનઃ પ્રાપ્ત કરવામાં આવે છે. પુનઃ પ્રાપ્તિ વાતાવરણથી મુક્ત પુનઃપ્રાપ્તિ પ્રશ્નની રચનામાં દર્શાવેલ, પદ્ધતિના ચાર મુખ્ય ઘટકો ધ્યાનમાં લેવામાં આવે છે.

(અ) વસ્તુઓ, પ્રલેખ અથવા પોતે જાતે નોંધો (જેમાં પ્રક્રિયા કરવાની હોય તે માહિતી ફાઈલોની સમગ્ર રચના કરવામાં આવે છે.)

(બ) માહિતી ઓળખો, પદો, નિર્દેશિકાના પદો, ચાવીરૂપ શબ્દો, લક્ષણો વગેરે (જેઓ નોંધો અથવા પ્રલેખોનું વર્ણન કરે છે અને ઉદાહરણમાં (નમૂનામાં) માહિતી વિષયવસ્તુનું નિરૂપણ પ્રતિનિધિત્વ કરે છે)

(ક) માહિતી વિનંતીઓ (જે પધ્ધતિમાં દાખલ થાય છે અને પુનઃપ્રાપ્તિ માટે સંગ્રહ કરેલ નોંધોની સાથે સરખામણી કરવામાં આવે છે.) અને

(૬) માહિતીની સુમેળતા (સંગ્રહિત માહિતીની બાબતો સાથે માહિતી વિનંતીનું જોડાણ ઘણીવાર પદ્ધતિના ઉપભોક્તા દ્વારા પૂરું પાડવામાં આવે છે)

માહિતી પુનઃ પ્રાપ્તિ એ એક એવી પદ્ધતિ છે જે માહિતીના સંગ્રહ, પુનઃપ્રાપ્તિ તથા સાર સંભાળા માટે સક્ષમ હોય છે. આ સંદર્ભમાં માહિતી પાઠ (સંખ્યાકિય અને તારીખ ડેટા સહિત), છાપ, દશ્ય, શ્રાવ્ય અને વિવિધ માધ્યમવાળી વસ્તુઓનો સમાવેશ કરે છે. આ રીતે માહિતી પુનઃપ્રાપ્તિ પદ્ધતિ સોફ્ટવેર કાર્યક્રમ ધરાવે છે જે ઉપભોક્તાને તેની જરૂરિયાતવાળી માહિતી શોધવામાં સુવિધા પૂરી પાડે છે. આ પધ્ધતિ માનાંકિત કમ્પ્યુટર કમ્પ્યુટર હાર્ડવેર અથવા વિશિષ્ટીકરણ થયેલ હાર્ડવેરનો ઉપયોગ કરી શકે છે. જે શોધના ઉપકાર્યો અને બિન પાઠવાળા સ્રોતને શોધ કરી શકાય તેવા માધ્યમમાં તબદીલ કરવામાં (દા.ત. માહિતીની પાઠમાં નકલ કરવી) સહાય પૂરી પાડે છે. જરૂરી માહિતી શોધવમાં ઉપભોક્તાના શિર ઉપર કેટલી સારી રીતે ઘટાડી શકાય છે. તેના માપ ઉપરથી માહિતી પદ્ધતિની સફળતા માપી શકાય છે. જરૂરી માહિતી ધરાવતા એકમના વાચનના તરફ દોરી જતા તમામ તબક્કાઓમાં ઉપભોક્તા કેટલો સમય વાપરે છે. તેને શિર ઉપર તરીકે રજૂ કરી શકાય. (દા.ત. શોધ પ્રશ્નનું નિર્માણ, શોધ કાર્યાન્વિત કરવી, વાચન માટે પસંદ કરવાની બાબતોનું શોધ પ્રશ્નના પરિણામોમાં પર્યવેક્ષણ કરવું, વિસંગત બાબતોનું વાચન.) માહિતી પધ્ધતિની સફળતા ઘણી વ્યક્તિલક્ષી છે અને શું માહિતી જરૂરી હતી ઉપભોક્તા સમગ્ર રીતે સ્વૈચ્છિક શું સ્વીકારે છે તેના ઉપર આધારિત છે. કેટલાક સંજોગોમાં જરૂરી માહિતીને 'તમામ માહિતી જે પદ્ધતિમાં છે તે ઉપભોક્તાની જરૂરિયાતને સુંસગત છે' મુજબ વ્યાખ્યાયિત કરી શકાય. આ રીતે શોધ રચના, શોધનો અમલ કરવો અને વિસંગત બાબતોનું વાચન, સમગ્ર માહિતી પુનઃપ્રાપ્તિના આ તમામ લક્ષણો છે.

સૌ પ્રથમ માહિતી પુનઃ પ્રાપ્તિ પધ્ધતિ ગ્રંથાલયોમાં માહિતીના સંગઠનની જરૂરિયાત શરૂ થઈ હતી. ગ્રંથાલય સૂચિઓ બાબતની ઓળખ અને પુનઃ પ્રાપ્તિ માટે સર્જન કરવામાં આવ્યું હતું. બાબત (Item) પદનો ઉપયોગ સંપૂર્ણ નાનો એકમ જેની પદ્ધતિ દ્વારા પ્રક્રિયા અને હસ્તક્ષેપ કરવામાં આવે છે તેની રજૂઆત માટે થાય છે. બાબત શબ્દની વ્યાખ્યા વિશિષ્ટ સ્રોત માહિતી સાથે કેવી રીતે કામ કરે છે તે પ્રમાણે જુદો પડે છે. સંપૂર્ણ પ્રલેખ જેવા કે પુસ્તક, સમાચાર પત્ર અથવા સામયિક બાબત હોઈ શકે છે. દરેક પ્રકરણ અથવા લેખને પણ બાબત તરીકે વ્યાખ્યાયિત કરી શકાય. જ્યારે સ્રોત વૈવિધ્યતા ધરાવે છે ત્યારે પદ્ધતિ સારાંશ જેવા કે પાઠ અથવા ફકરાના નાના ભાગ જેવી બાબતોને નાના સ્તર સુધી વર્ણવી શકાય તેવી જટિલ પ્રક્રિયાનો સમાવેશ કરે છે.

સક્ષમ પ્રક્રિયા કરવાની ક્ષમતાવાળા સસ્તા અને વધુ ઝડપી, અને દ્વિતીય કક્ષાના બિન ઉત્પાદનોનું વિશાળ માત્રામાં સંગ્રહ ક્ષમતા ધરાવતા પર્સનલ કમ્પ્યુટર પધ્ધતિઓના આગમનને કારણે સામાન્ય ઉપભોક્તાને માટે વિશાળ પાઠ આધારિત માહિતી ડેટા સંપુટો વ્યાપારી ધોરણે પૂરા પાડવાનું સંભવિત બનેલ છે.

માહિતી પુનઃ પ્રાપ્તિમાં 'સુસંગત' પદ જરૂરી માહિતી ધરાવતી બાબત દર્શાવવા ઉપયોગ થાય છે ઉપભોક્તાને અનુલક્ષીને 'સુસંગત' અને ''જરૂરી' સમાનાર્થી શબ્દો છે. પધ્ધતિને અનુલક્ષીને માહિતી, શોધ વિધાન (અર્થાત્ શોધ વિધાનની સરખામણીના ધોરણ)ને સુંસગત હોઈ શકે છે. જો કે તે ઉપભોક્તાને સુસંગત (દા.ત. ઉક્ભોક્તા અગાઉથી જ માહિતી જાણી ચૂકેલ હોય) તે જરૂરી નથી.

ઘણીવાર ઉપભોક્તા જે વિષય ક્ષેત્રમાં શોધ કરવાની હોય તેનો નિષ્ણાત હોતો નથી અને વિશિષ્ટ નિર્ધારિત શબ્દભંડોળ જે ખાસ તે વિષયક્ષેત્રમાં અનન્ય હોય છે તેની ઉણપ અનુભવે છે. આ બાબત શોધપ્રશ્ન નિર્માણની જટિલતામાં વધારો કરે છે. ઉપભોક્તા શોધ પ્રશ્નની શરૂઆત જરૂરિયાતવાળી માહિતીની સામાન્ય વિભાવનાથી કરે છે પરંતુ તેને ચોક્કસ શું જોઈએ છે. તેની વ્યાખ્યા ખાસ વિષય વિસ્તાર સંબંધિત શબ્દભંડોળ અંગેનું મર્યાદિત જ્ઞાનની સાથે સાથે ચોક્કસ કઈ માહિતી જોઈએ છે તે તરફ ધ્યાન કેન્દ્રિત કરવાની ઉણપ, ઉપભોક્તાને ચોકસાઈરહિત અને કેટલીકવાર ગેરમાર્ગે દોરતા પદો તરફ લઈ જાય છે. ઉપભોક્તા જે વિષયમાં શોધ કરવાની હોય તે વિષયનો નિષ્ણાત હોય તો પણ, લેખકના શબ્દભંડોળના જ્ઞાનની ઉણપ તેને ચોક્કસ પદો પસંદ કરવાની ક્ષમતામાં અવરોધ ઊભો કરે છે. કેટલાક માહિતી અનુલક્ષીને કુદરતી અવરોધો હોય છે જેવા કે ઉપભોક્તાને શુ માહિતીની જરૂરિયાત ભાષાઓ અર્થની અરૂઢ થયેલ સંદિગ્ધતામાંથી આવે છે, ઉપભોક્તાને શું માહિતી જોઈએ છે તેને રજૂ કરવાની મર્યાદાઓ અને ઉપભોક્તાના શબ્દભંડોળ રાશિ અને ડેટા સંપુટોમાં તે બાબતો

અગે લેખકોના શબ્દભંડોળ વિવિધ અર્થ, ટૂંકાક્ષરી શબ્દો જેમાં એક શબ્દના જુદા જુદા થતા હોય તેવી ખામીઓ અનુભવે છે. ઘણા ઉપભોક્તાને સારા શોધ વિધાનની રચનામાં મુશ્કેલીઓ હોય છે. જટિલ ઉપભોક્તાને તેની સાથે નોંધપાત્ર અનુભવ નથી હોતો. વળી, બુલીયન તર્ક વિધાન સાથે યોગ્ય શક્તિ પણ નથી હોતી, બુલીયન તર્કશાસ્ત્રનો ઉપયોગ, ડેટા સંપુટ સંચાલન પદ્ધતિઓના વિકાસથી જ ઉત્તરાધિકાર ધરાવે છે. વિવિધ માધ્યમ પણ શોધ વિશિષ્ટકરણમાં જટિલતાના સ્તરમાં વધારો કરે છે.

આ રીતે માહિતી પુનઃપ્રાપ્તિ પદ્ધતિએ શોધ વિશિષ્ટિકરણના પ્રશ્નોના નિવારણ માટે સાધનો (પ્રયુક્તિઓ) પૂરી પાડવી જોઈએ. ખાસ કરીને શોધ માટેના સાધનો (પ્રયુક્તિઓ) એ ઉપભોક્તાને સ્વયંસંચાલન અને પદ્ધતિ આંતરક્રિયા દ્વારા શોધ વિશિષ્ટીકરણ જે ઉપભોક્તાની જરૂરિયાત અને લેખકોની લખવાની વૈવિધ્યતાનું નિરૂપણ કરે છે તેમાં મદદ કરવી જોઈએ

કોમ્પ્યુટર આધારિત આધુનિક માહિતી પુનઃપ્રાપ્તિ પદ્ધતિઓના મુખ્ય પાયાગત ત્રણ અંગો હોય છે. તે આ પ્રમાણે છે. ડેટા સંપુટો, માહિતી મેળવનાર અને પુનઃપ્રાપ્તિ પ્રયુક્તિઓ, સાધનો, નમૂનાઓ (Models) અને પ્રક્રિયાઓ જે માહિતી શોધક અને ડેટા સંપુટ વચ્ચેના અવકાશમાં પુલ બાંધ પ્રયત્ન કરે છે. (બંનેને સાથે સાંકળી લેવાનો પ્રયત્ન કરે છે.)

18.2.1 ડેટા સંપુટ (Database)

તેની રચના માહિતી અને વિશિષ્ટ ઉપભોક્તા સમૂહોને માહિતી સેવા પૂરી પાડવાના હેતુથી પ્રાપ્ત કરેલ માહિતી સ્ત્રોતનો સમાવેશ કરીને કરવામાં આવે છે. માહિતી અને અન્ય સ્રોતની પસંદગી તેમને સેવા આપવાના આ હેતુ આધારિત હોય છે. ડેટા સંપુટ પાસે તાર્કિક અને ભૌતિક સંગઠન હોય છે. તેની ગોઠવણી ઉપભોક્તાના અભિગમ આધારિત હોય છે. વાઙ્મયસૂચિગત ડેટા સંપુટોનો મુખ્ય હેતુ પ્રમાણિત વાઙ્મયસૂચિગત માહિતી એકત્રિત કરવાનો અને પાસ પાસે રાખવાનો, તથા ટેકનીકલ પેપર્સ અને મોનોગ્રાફ માટે અને નિર્દેશિકા પદો પ્રદાન કરવાનો છે. સંપૂર્ણ પાઠ ડેટા સંપુટોના ઉદાહરણમાં, પ્રલેખનો સંપૂર્ણ પાઠ તથા તેની સાથે વાઙ્મયસૂચિગત, માહિતી ઊંડાણપૂર્વક ઉપલબ્ધ કરવામાં આવે છે.

માહિતી પુનઃ પ્રાપ્તિ પદ્ધતિઓ ચોક્કસ માળખા સિવાય માહિતી સંગ્રહના મુક્ત સાધનો ઉપયોગને કરી વધી રહી છે. પ્રણાલિકા મુજબ કાગળ ઉપર નોંધો રાખવાની જેમ દા.ત. વ્યક્તિના નામ માટે અલગ લાઈન, સરનામા માટે અલગ લાઈન, ટેલિફોન નંબર માટે અંકોવાળી જગ્યા અને તેવી બીજી માહિતી માટે નોંધો રાખવાની હોઈ શકે છે. કમ્પ્યુટર પધ્ધતિઓમાં હવે સ્થાનની ચોકસાઈ ઓછી દર્શાવી, માહિતી સંગ્રહ કરવાનું સામાન્ય બની ગયું છે જે વિશિષ્ટ લક્ષણો જેવા કે નામ અથવા સરનામું પ્રાપ્ત કરવું મુશ્કેલ બનાવે છે. વધતો જતો વર્ડ પ્રોસેસીંગનો ઉપયોગ, ડેસ્કટોપ પર્સનલ કમ્પ્યુટરનું વધતું જતું ઉત્પાદન અને ઓપ્ટિકલ સ્કેનર જે સામગ્રીનું વાચન કરી સંગઠિત ડેટા સંપુટમાં બદલી શકે છે જે એવી માહિતી સંગ્રહ કરે છે કે જેનું માળખાયુક્ત પુનઃ સર્જન કરવું વધુ મુશ્કેલ બનાવે છે, આ બાબત માહિતી સંગ્રહ પદ્ધતિઓને વધુ લોકપ્રિય બનાવે છે.

બીજી બાજુ મુક્ત સ્વરૂપમાં, બિન માળખાયુક્ત રીતે સંગ્રહ કરવાને કારણે ચોકસાઈ, વિજાણુ સાધનો જે માહિતી પુનઃપ્રાપ્તિ અને તેમાં હસ્તક્ષેપ જે રીતે શક્ય બનાવે છે તે કાગળ આધારિત પદ્ધતિમાં શક્ય નથી, માહિતી સંગ્રહ અને પુનઃ પ્રાપ્તિના વિજાણુયુક્ત સાધનો ઉપર વધતો જતો વિશ્વાસ તેને વધુ સ્પષ્ટ બનાવે છે. માહિતી પુનઃપ્રાપ્તિમાં શક્તિ અને લચીલાપણાનો તફાવત મુદ્રિત નિર્દેશિકાઓ અને તેના વિજાણુ સ્વરૂપમાં માહિતી શોધના ઉદાહરણ દ્વારા માપી શકાય છે. મુદ્રિત નિર્દેશિકામાં માહિતી શોધ, ઉપભોક્તાને પસંદ કરેલ ચાવીરૂપ શબ્દો અથવા નિર્દેશિકા પદો જેને પ્રમાણમાં મર્યાદિત રીતે જોડાણ કરવા દે છે. વિજાણુ ડેટા સંપુટ સામાન્ય રીતે ઉપભોક્તાને સ્વતંત્રતા આપે છે અને પદો વચ્ચેનો વિષયવંશાનુક્રમ અને સામીપ્ય સંબંધ વધુ ચોકસાઈથી વ્યાખ્યાયિત કરવા દે છે.

18.2.2 માહિતી આધાર (Information Base)

માહિતી પુનઃ પ્રાપ્તિ પધ્ધતિ, ઉપભોક્તા પદ્ધતિમાં રજૂ કરેલ શોધ પ્રશ્નને આધારે માહિતીનું સ્થાન દર્શાવે છે અને રજૂ કરે છે. શોધપ્રશ્ન પુનઃપ્રાપ્તિ કરવાની હોય તે માહિતીના ચોકસાઈયુક્ત લક્ષણો દર્શાવી શકે છે અથવા પુનઃપ્રાપ્તિ કરવાની હોય તેવી માહિતીના લક્ષણોના આરંભિક અનુમાન તરીકે ભાગ્યે જ દર્શાવતી હોય તેવી જરૂરી માહિતીની અંદાજિત રજૂઆત કરી શકે છે. ચોકસાઈયુક્ત શોધ પ્રશ્ન સામાન્ય રીતે જેમાં જવાબ એક જ જરૂરી હકીકત પ્રશ્ન સંદર્ભમાં હોય દા.ત. 'ગુજરાતની રાજધાની કઈ છે' તેવા સ્વરૂપમાં જોવા મળે છે. જે પ્રશ્નમાં જવાબો વધુ જટિલ અથવા ગૂંચવણ ભરેલા હોય તેવા શોધ પ્રશ્નોની રજૂઆતમાં ચોકસાઈ ઓછી હોય છે. એક ઉદાહરણ લઈએ. એક સંસ્થા બહુભાષી પ્રલેખોનો સંગ્રહ અને પુનઃપ્રાપ્તિ કરતી હોય તેવી પધ્ધતિ આંતરિક રીતે ભાડે રાખી પ્રસ્થાપિત કરવા ઈચ્છે છે. કર્મચારી સંચાલન વિભાગના વડા (મંત્રી) 'બહુભાષી પદ્ધતિઓનો કોને અનુભવ છે' એવો પ્રશ્ન ઉઠાવી શકે. આ પ્રશ્ન માટે અપેક્ષિત જવાબો અનેક છે. માહિતી પુનઃ પ્રાપ્તિ જુદા જુદા પ્રકારના પ્રશ્નો મેળવી શકે, જુદી જુદી રીતે જવાબો આપી શકે, વ્યક્તિગત હકીકતો અથવા અપેક્ષિત સુસંગત બાબતોના સમૂહની પુનઃપ્રાપ્તિ માટે સક્ષમ હોવી જોઈએ.

આ પ્રકારનું લચીલાપણું (Flexibility) સ્રોતની વિશાળતા અને જરૂરી માહિતીમાં વૈવિધ્યતાને કારણે જરૂરી છે. જેના પાયામાં બેલ્કિન, બ્રુક્સ અને ઓડી (1972) જેનો સંદર્ભ આપે છે તે 'જ્ઞાનની સ્થિતિમાં વિસંગતતા રહેલી છે. (ASK-Anomalous state of knowladge). આ જ્ઞાનની સ્થિતિની વિસંગતતાઓ, માહિતી મેળવનારના જ્ઞાનના આધારની ખાસ હકીકતની રજૂઆતમાં ઉણપ હોઈ શકે છે, તેઓ ખૂટતી માહિતીનો વિશાળ વ્યાપ રજૂ કરી શકે છે અથવા જ્ઞાનના કે બંધારણની ઉણપ રજૂ કરી શકે છે. ઘણીવાર નાની જરૂરિયાત એક હકીકત દ્વારા જવાબ આપી શકાય છે જ્યારે ઘણી મોટી જરૂરિયાત તવા કોઈ વ્યક્તિના જ્ઞાનના વધુ બિન બંધારણીય વિસ્તાર આધારમાં અપેક્ષિત જવાબો વિવિધ રીતે, અનેક હકીકતો અને પ્રલેખો દ્વારા આપી શકાય છે. માહિતી સંપુટ તે ડેટા સંપુટો અથવા સંગ્રહો બેમાંથી જે હોય તે, તેનો ઉપયોગ કરનાર લોકોની અપેક્ષિત જરૂરિયાત પહોંચી વળવા સંગઠિત અને બંધારણયુક્ત હોવો જોઈએ, અથવા જરૂરી માહિતી માટે શું જરૂરી છે, લોકો કેવી રીતે સમજી શકશે અને પધ્ધતિનો કેવી રીતે ઉપયોગ કરી શકે તેનું સર્જન કરવું જોઈએ.'

18.2.3 માળખાગત શોધ ભાષા (Structured Query Language (SQL))

માળખાગત શોધ ભાષા (SQL - Structured Query Language) ભાષાએ ડેટા સંપુટની માહિતી પ્રાપ્તિ અથવા માહિતી સુધારણા માટે ઉપયોગમાં લેવામાં આવતી શોધપ્રશ્ન ભાષા છે. SQL કેટલાક નિયંત્રકો 'દાખલ કરો (Insert), 'સુધારો' (Update), રદ કરો (Delete) નો સમાવેશ કરે છે. શોધ પ્રશ્ન નિયંત્રક ભાષાનું સ્વરૂપ લે છે તે ઉપભોક્તાને ડેટા પસંદ કરવા, દાખલ કરવા, સુધારવા, સ્થાન શોધી કાઢવા અને તેવા બીજા કાર્યો તરફ લઈ જાય છે. કાર્યક્રમની આંતરસન્મુખતા પણ હોય છે. આઈબીએમ (IBM) દ્વારા સૌ પ્રથમ 1975 માં ભાષાનું સર્જન કરવામાં આવ્યું હતું અને તેને SEQUEL - બંધારણીય અંગ્રેજી ભાષા શોધ પ્રશ્ન (Structured English Query Language) તરીકે ઓળખવામાં આવતી હતી. ત્યારબાદ તેમાં ઓરેકલ કોર્પોરેશનના અનેક પ્રભાવ સાથે તેમાં અનેક સુધારા થવા પામેલ છે. આજે SQL પ્રમાણિત ભાષા તરીકે સ્વીકારવામાં આવે છે પરંતુ તેના ઘણાં સ્વરૂપો જેવા કે mSQL, mySQL છે. ઘણા ડેટા સંપુટ ઉત્પાદનો જેવા કે MS-Access, SQL server and Oracle support, SQL , પ્રમાણિત ભાષા સાથેના માલિકીપણાના વિસ્તરણો છે. કેટલીક માહિતી પુનઃપ્રાપ્તિ

પદ્ધતિઓમાં શોધ પ્રશ્ન દ્વારા નક્કી કરાયેલ લક્ષણો ધારણ કરેલ હોય એવી હકીકતો અથવા પ્રલેખોની જ શોધ કરવાની મર્યાદા હોય છે. આવી પધ્ધતિઓ ઘણીવાર ડેટા સંપુટ પધ્ધતિઓ અથવા ડેટા સંપુટ SQL તરીકે અથવા પદ્ધતિમાં સ્વીકાર્ય હોય તેવા માન્ય વિવિધ સ્વરૂપો તરીકે ઓળખાય છે. SQL પુનઃપ્રાપ્તિ માટેના પદોના મૂલ્ય માટેના ચોકસાઈથી સ્પષ્ટ કરેલ લક્ષણોને દાખલ કરવા દે છે.

SQL શોધ પ્રશ્નનું ઉદાહરણ આ પ્રમાણે હોઈ શકે :

અભ્યાસક્રમોમાંથી વિદ્યાર્થીનું નામ પસંદ કરો, અંજલિ કપૂર અને વિભાગ સંચાલન વિજ્ઞાન.

◆ **તમારી પ્રગતિ ચકાસો :**

1. માહિતી પુનઃપ્રાપ્તિ એટલે શું ?

નોંધ : (1) તમારો જવાબ નીચે આપેલ જગ્યામાં લખો.

(2) તમારો જવાબ આ એકમને અંતે આપેલ જવાબો સાથે મેળવો.

....

....

....

....

....

....

18.3 માહિતી પુનઃપ્રાપ્તિની તક્નિકો (INFORMATION RETRIEVAL TECHIQUES)

માહતીનું સંગઠન અને માહિતી પુનઃપ્રાપ્તિની વિવિધ તક્નિકોની વિકાસ એ સંશોધનનું મુખ્ય ક્ષેત્ર રહેવા પામેલ છે. કમ્પ્યૂટર ટેકનોલોજીમાં વિકાસની સાથે કમ્પ્યૂટર આધારિત માહિતી પુનઃપ્રાપ્તિ પદ્ધતિઓ ઉપર વધુ ભાર આપી આ ક્ષેત્રમાં રસ વધુ તાજો થવા પામેલ છે.

જરૂરી તક્નિકોની મુખ્યત્વે ચોકસાઈયુક્ત મેળવણી, ઉત્તમ મેળવણી તથા અંશતઃ મેળવણી કક્ષાઓ પાડી શકીએ છીએ અને ત્યાર પછી વ્યક્તિગત અથવા નેટવર્ક તક્નિકોમાં વિભાજિત કરી શકીએ છીએ. આ તક્નિકોને ફરીથી આગળ વિશિષ્ટ તક્નિકો જેવી કે જૂથ, સંભાવના, સદિશ અને તેવી બીજી તક્નિકોને અનુકૂળ હોય તે મુજબ વિભાજિત કરી શકાય છે. ચોક્કસ મેળવણી પુનઃપ્રાપ્તિ એ સૌથી વ્યાપક ઉપયોગમાં લેવાતી એક તક્નિક છે. તે બુલીયન સંપૂર્ણ પાઠ અથવા ક્રમિક શોધ (String Search) દ્વારા અમલીકરણ કરાય છે. તેનો લાભ છે કે તે સમજવામાં અને ઉપયોગમાં સરળ છે અને મોટાભાગની પદ્ધતિઓમાં પ્રાપ્ય છે પરંતુ તેના ગેરલાભ ઘણા છે. તે ઘણા સુસંગત પાઠ ગુમાવે છે જે શોધ પ્રશ્નની ફક્ત અંશતઃ મેળવણી કરે છે. તે પ્રલેખનો ક્રમ આપતી નથી. આપણે શોધ સમયે સંશોધકની અગ્રતા માટે અથવા સંશોધન પ્રશ્નની શરૂઆત માટે ફક્ત આ તક્નિકના ઉપયોગથી પુનઃપ્રાપ્તિ થતી માહિતીથી સંતોષ થશે તે અંગે વિચારવું પડે છે. વિવિધ પરિમાણીય અને વિવિધ વિદ્યાશાખા ધરાવતા અને કેટલીક બાહ્ય વિસ્તારો આવરી લેતા સંશોધનોમાં સંભવિત સક્રિય પ્રસાર કરવાની તક્નિક (spreading activation technique) અથવા ઉદ્ધૃત નોંધોનો વ્યુહ જેમાં પાયાગત પ્રલેખમાં સૌથી વધુ વાર ઉપયોગ હોય તેનો ઉપયોગ આ ઘટનામાં વધુ અસરકારક થશે. ઉત્તમ મેળવણી, બે અથવા વધુની તુલનાનો સંદર્ભ આપે છે અને ચોકસાઈયુક્ત મેળવણી જેવી એક જ સરખી ન હોય. વધુ ઊંડાણપૂર્વક અભ્યાસ માટે તમે એકમ નં. 17 વાંચી શકો છો.

પુનઃપ્રાપ્તિ માટે શોધ પ્રશ્નોની મેળવણીનો પ્રશ્ન અભ્યાસનું બીજુ પાસું છે. અહીં પ્રશ્ન એ છે કે વિશિષ્ટ પુનઃપ્રાપ્તિ તક્નિકો છે અથવા ઘણા પ્રશ્નો માટે એક જ પુનઃપ્રાપ્તિ તક્નિકનો આપણને વિકલ્પ મળી શકે છે.

આમ છતાં બૌદ્ધિક માહિતી પદ્ધતિઓના સંદર્ભમાં ઉપભોક્તાઓની જુદી જુદી પરિસ્થિતિઓ માટે જુદી જુદી પુનઃપ્રાપ્તિ તક્નિકો જરૂરી છે. જ્ઞાનનું આલેખન અને અન્ય પુનઃપ્રાપ્તિ તક્નિકો સમગ્ર સંશોધન પ્રક્રિયાને સંકલિત અભિગમ પૂરો પાડે છે.

◆ **તમારી પ્રગતિ ચકાસો**

(2) જુદી જુદી માહિતી પુનઃપ્રાપ્તિ તક્નિકો કઈ કઈ છે ?

નોંધ : (1) તમારો જવાબ નીચે આપેલ જગ્યામાં લખો

(2) તમારો જવાબ આ એકમને અંતે આપેલ જવાબો સાથે મેળવો.

....

....

....

....

....

....

18.4 મોડેલ આધારિત નિવેશ/નિર્ગમન (MODELS BASED ON INPUT/OUTPUT)

નિવેશ અને નિર્ગમનના આધારે માહિતી પુનઃપ્રાપ્તિના નમૂનાઓના મૂળભૂત ત્રણ કક્ષાના જૂથો પાડી શકાય છે.

(1) ડેટા પુનઃપ્રાપ્તિ નમૂનો

(2) માહિતી પુનઃપ્રાપ્તિ નમૂનો

(3) જ્ઞાન પુનઃપ્રાપ્તિ નમૂનો

18.4.1 ડેટા પુનઃ પ્રાપ્તિના મોડેલ (Data Retrieval Model)

ડેટા પુનઃપ્રાપ્તિ નમૂનો આવશ્યક હોય તેવા ડેટા હસ્તગત કરે છે. આપણી સમજના હેતુ માટે, ડેટા તરીકે પ્રક્રિયા ન કરેલ હોય તેવી માહિતી અથવા પ્રાથમિક તબક્કાની માહિતી લઈ શકાય. ડેટા પૂર્વગ્રહ રહિત હકીકત છે જેનો ઉપયોગ માહિતીની રચના માટે થાય છે. દા.ત. આપણે કહી શકીએ કે જયપુર શહેરની વસ્તી એંસી (80) લાખ છે. આ ડેટા છે. આ રીતે વસ્તી માહિતી પધ્ધતિએ ડેટા પુનઃપ્રાપ્તિ પદ્ધતિ છે. તે જ રીતે નેશનલ સેમ્પલ સર્વે ઓર્ગેનાઈઝેશન અને સેન્ટ્રલ સ્ટેટેસ્ટીકલ ઓર્ગેનાઈઝેશનને સંખ્યાત્મક ડેટા પદ્ધતિઓ તરીકે લઈ શકાય. ડેટા પુનઃપ્રાપ્તિ નમૂનાને વિવિધ ધોરણ જેવા કે ગુણધર્મો, જૂથ અને અન્ય જુદા જુદા એકમો આધારિત સંગઠિત માળખા તરીકે ઓળખવામાં આવે છે. આ બાબતોની નામાંકનયુક્ત રજૂઆત માટે જરૂરિયાત છે. આ પ્રકારની નામાંકનયુક્ત રજૂઆત માટે અન્ય પ્રકારોનાં સંગઠનમાંથી પણ પ્રાપ્ત હોવા જોઈએ. ડેટાનો શોધક વિશિષ્ટ માહિતી પુનઃપ્રાપ્તિ માટે આવે છે તેથી માહિતીની રજૂઆતમાં ખૂબ ચોકસાઈ આવશ્યક છે તેથી ડેટા પુનઃપ્રાપ્તિ નમૂનો એ માહિતી પુનઃપ્રાપ્તિ છે.

જેમાં સરખામણી માટે વિશિષ્ટ તક્નિકો જરૂરી છે. દા.ત. સમાવેશ થતા એકમોનું નામાંકનયુક્ત બંધારણ માળખું અને તેના ગુણધર્મો.

18.4.2 માહિતી પુનઃ પ્રાપ્તિના મોડેલ (Information Retrieval Model)

માહિતી એક હેતુલક્ષી ડેટા છે. તે વાસ્તવિક રીતે સંબંધયુક્ત માળખામાં કેટલાક ડેટાને જોડે છે. તેથી માહિતી પુનઃપ્રાપ્તિ એ વધુ જટિલ મોડેલ છે. સામાન્ય રીતે વિવિધ પરિમાણીય સંબંધનો તેમાં સમાવેશ કરવો પડે છે. તે નામાંકનયુક્ત માળખાની જેમ સરળતાથી આજ્ઞાંકિત બનાવી શકાતું નથી. માહિતીની રજૂઆત ગણિતશાસ્ત્રના કેટલાંક જોડાણોના ઉપયોગ દ્વારા સંબંધાત્મક ડેટા આધારિત માળખા ઉપર કરવામાં આવે છે. જરૂરી માહિતીની રજૂઆત પણ જટિલ અને સમયનો વ્યય કરનારી છે. તે લાંબા સમય માટે માહિતી પુનઃપ્રાપ્તિ નમૂનામાં આ પ્રકારની સુવિધા અને આંતરસન્મુખતા પ્રયોજવી જોઈએ.

18.4.3 જ્ઞાન પુનઃ પ્રાપ્તિના મોડેલ (Knowladge Retrieval Model)

જ્ઞાન એ સામાન્ય પ્રકારની માહિતીનું એક સંગઠન છે તે સામાન્ય રીતે માનવના મગજમાં આકાર પામે છે. માનવ મગજ અનુમાન કરે છે અને તેણે મેળવેલ માહિતીની સાથે કેટલાક સમન્વય કરી સંકલન કરે છે તેથી જ્ઞાન એ માહિતીનું વિલિનીકરણ છે. નિર્ણય પ્રક્રિયામાં અને પ્રશ્નોના ઉકેલ મેળવવામાં સહાય પ્રદાન કરવા બુદ્ધિજન્ય જ્ઞાન આધારિત માહિતી પુનઃપ્રાપ્તિના નમૂના બહાર આવવા લાગ્યા છે. આવી પદ્ધતિઓ ત્રણ બાબતો ધરાવતી હોય છે.

(અ) જેને જ્ઞાનનો આધાર કહી શકાય અથવા માહિતીનું જ્ઞાનમાં પરિવર્તન કરવા માટેના નિયમોના કુલ સમૂહનો સંગ્રહ. તે જ્ઞાન પ્રાપ્તિ પધ્ધતિને પણ પ્રયોજિત કરે છે.

(બ) પદ્ધતિની બીજી બાબત અનુમાન યંત્ર છે. અનુમાન યંત્ર સંયોજિત જ્ઞાનની પ્રાપ્તિ માટે નિયમોના સમન્વયમાંથી યોગ્ય માહિતી મેળવવા સક્ષમ હોય છે. આ મેળવવાની પ્રક્રિયા, સંખ્યાત્મક અને બિન - સંખ્યાત્મક તક્નિકોનો ઉપયોગ કરી અનુમાનિક તર્કશાસ્ત્ર ઉપર આધારિત હોય છે.

(ક) ઉપભોક્તા આંતરસન્મુખતા અર્થાત્ નમૂનામાં પ્રત્યાયન પ્રક્રિયા જે પ્રત્યાયન કરી શકે તેવા સ્વરૂપે માહિતી મેળવી આંતરક્રિયાના હેતુઓ માટે તેનું ડેટા સંપુટ સંકેતોમાં પરિવર્તન કરવા સક્ષમ છે. આ રીતે જ્ઞાન પુનઃપ્રાપ્તિ નમૂનો માહિતીની પ્રક્રિયા, સંગઠન અને પુનઃપ્રાપ્તિ માટે સુઘડ નમૂનો છે.

◆ **તમારી પ્રગતિ ચકાસો**

(3) પુનઃપ્રાપ્તિના વિવિધ નમૂનાઓની ટૂંકમાં ચર્ચા કરો

નોંધ : 1. તમારો જવાબ નીચે આપેલ જગ્યામાં લખો

2. તમારો જવાબ આ એકમને અંતે આપેલ જવાબો સાથે મેળવો.

....

....

....

....

....

....

18.5 મોડેલ આધારિત સિદ્ધાંતો અને સાધનો (MODELS BASED ON THEORIES AND TOOLS)

અન્ય વિદ્યાશાખામાં પ્રાપ્ય સિદ્ધાંતો અને પદ્ધતિઓ/સાધનોને આધારે માહિતી પુનઃપ્રાપ્તિના પ્રશ્નોના સંતોષજનક ઉકેલ મેળવવા સંખ્યાબંધ નમૂનાઓ વિકસાવવામાં આવ્યા છે.

18.5.1 બુલીયન પુનઃ પ્રાપ્તિ મોડેલ (Boolean Retrieval Model)

બુલીયન તર્કશાસ્ત્ર ઉપભોક્તાને જે માહિતી જરૂરી હોય તેને વ્યાખ્યાયિત કરવા બહુવિધ વિભાવનાઓ તાર્કિક રીતે જોડવા અનુમતિ આપે છે. જટિલ રીતે બુલીયન કાર્યો, કોઈ એક બાબતમાં ગમે ત્યાં હોય તેવા લક્ષણોને ઓળખી કાઢી તેની પ્રક્રિયા કરે છે. AND, OR and NOT જટિલ બુલીયનાં કારકો છે. આ ક્રિયાઓ બે વિભાગો કરવા, બેનું જોડાણ કરવા અને ભિન્ન ભિન્ન કાર્યવાહીનો ઉપયોગ કરી પ્રાયોજિત કરવામાં આવે છે. કેટલીક પદ્ધતિઓ 'Exclusive OR' ની વિભાવના દાખલ કરે છે પરંતુ તે અન્ય કારકોના ઉપયોગની તુલનામાં સાધારણ જટિલ છે અને તે ઉપભોક્તાઓને સામાન્ય રીતે ઉપયોગી નથી કારણ કે મોટાભાગના ઉપભોક્તા તે સમજી શકતા નથી. સામાન્ય રીતે બુલીયન કારકો નીચેના પરિણામો પૂરા પાડે છે.

'A AND B'	બંને પદ એ અને બી હોય તેવી બાબતોની પુનઃપ્રાપ્તિ કરે છે.
'A OR B'	પદ એ અથવા પદ બી અથવા બંને પદ હોય તેવી બાબતોની પુનઃપ્રાપ્તિ આપે છે.
'A NOT B'	પદ એ હોય અને પદ બી ન હોય તેવી બાબતોની પુનઃપ્રાપ્તિ કરી આપે છે.

18.5.1.2 ભારિત બુલીયન પુનઃપ્રાપ્તિ નમૂના (Weight Boolean Retrieval Model)

શોધ પ્રશ્ન તૈયાર કરવાના મુખ્ય બે અભિગમ બુલીયન અને કુદરતી ભાષા છે. કુદરતી ભાષા શોધ પ્રશ્નો આંકડાકીય નમૂનામાં સરળતાથી રજૂ કરાય છે અને સામ્યતાના માનદંડ દ્વારા ઉપયોગ કરી શકાય છે. જ્યારે બુલીયન શોધ પ્રશ્નો ભારિત નિર્દેશિકા પદ્ધતિઓ સાથે જોડાયેલ હોય ત્યારે પ્રશ્નો ઉદ્ભવે છે. કેટલાક પ્રશ્નો તાર્કિક કારકો (AND, OR, NOT) ભારિત મૂલ્યો સાથે કેવી રીતે કાર્ય કરે છે અને ભાર શોધ પ્રશ્ન પદ સાથે કેવી રીતે જોડાયેલ છે તેની સાથે જોડાયેલ હોય છે. જો કારકોનું અર્થઘટન તેમના સામાન્ય અર્થઘટનાત્મક સ્વરૂપે કરવામાં આવે તો તેઓ ઘણા મર્યાદિત અથવા સામાન્ય રીતે કાર્ય કરે છે. સાલ્ટોને (1979) કારકોનો ચોક્કસ વ્યાખ્યામાં ઉપયોગ કરીને બતાવ્યું છે. ઉપભોક્તા દ્વારા અપેક્ષિત પુનઃપ્રાપ્તિ ગૌણ આશાસ્પદ રહે છે. ચોક્કસ વ્યાખ્યાથી નજીક સંબંધમાં પ્રશ્ન ક્રમનો છે જે શુદ્ધ બુલીયન પ્રક્રિયામાં જોવા મળતો નથી. સાલ્ટોને નિર્દેશિકાઓમાં ભાર પ્રાપ્ય નથી હોતો એવી ધારણા સાથે બુલીયન શોધ પ્રશ્નો અને ભારિત શોધ પદોને ભેગા કરવાના પ્રશ્નોમાં વધારાની આંતરદૃષ્ટિ ઉમેરી તેનો હેતુ સામાન્ય બુલીયન કારકોને અમલમાં મૂકવા અને ત્યારબાદ ભારીકરણ તક્નિકોના ઉપયોગથી પરિણામો સુધારવાનો હતો.

નિર્દેશિકા પદોનું ભારીકરણ કરવું એ પ્રણાલિકાગત નિર્દેશીકરણ પદ્ધતિઓમાં સામાન્ય નથી. ભારીકરણ એ એકમમાં ઉપયોગમાં લેવાયેલ નિર્દેશિકા પદનું મહત્વ પ્રદાન કરવાની પ્રક્રિયા છે. ભારે નિર્દેશિકા પદ જે એકમમાં રજૂ થવા પામેલ છે તેની સાથે જોડાયેલ વિભાવનાની માત્રા પ્રદર્શીત કરવી જોઈએ ત્યારે ડેટા સંપુટના એકમોમાં વિભાવના કેટલે સુધી વર્ણવેલ છે તે જુદા પાડવામાં મદદ કરવી જોઈએ. ભાર પ્રદાન કરવાની પ્રણાલિકાગત પ્રક્રિયા નિર્દેશિકરણ કર્તા શિર ઉપર વધારો કરે છે (કાર્યમાં વધારો કરે છે.) અને ભારના સંગ્રહ માટે જટિલ ડેટા માળખાની જરૂરિયાત રહે છે. ભારિત નિર્દેશીકરણ પદ્ધતિઓમાં નિર્દેશિકા પદ જેની સાથે જોડાયેલ પ્રલેખની વિભાવનાની રજૂઆતનું મૂલ્ય મૂકવાનો પ્રયત્ન કરવામાં આવ્યો છે. નિર્દેશિકા પદોનું મૂલ્ય શૂન્ય અને એકની વચ્ચે સામાન્યીકરણ કરવામાં આવે છે. ઊંચો ભાર, પદની વિભાવનાની એકમમાં વધુ ચર્ચા દર્શાવે છે. ભાર ડેટા સંપુટમાં એકમોની સંખ્યા જે તેવી જ વિભાવના ધરાવતા હોય તેના માટેના હિસાબમાં વ્યવસ્થિત કરી શકાય છે.

શોધ પ્રશ્ન એકમ શોધ પ્રશ્નને સંતોષશે તેનું લગભગ અનુમાન કરવા ઉપયોગમાં લેવાયેલ પ્રમાણ મૂલ્ય (મૂલ્ય ક્રમ) નક્કી કરવા શોધ પ્રશ્નના પદોને પ્રદાન કરેલ અન્ય ભાર સાથે ભાર પ્રક્રિયા કરવામાં આવે છે. પરિણામો ઉપભોક્તાને સૌથી વધુ સંખ્યાથી સૌથી ઓછી સંખ્યાવાળા ક્રમ મૂલ્ય મુજબ દર્શાવવામાં આવે છે.

જો પદોને પ્રદાન કરેલ ભાર 0.0 થી 1.0 ની વચ્ચે હોય તો, ઉપભોક્તા દરેક પદને નોંધપાત્ર સાર્થકતા મુજબ અર્થઘટન કરે છે. મૂલ્ય 1.0 એ બુલીયન શોધપ્રશ્નનું ચોક્કસ અર્થઘટન છે એમ ધારવામાં આવે છે. મૂલ્ય 0.0 નું અર્થઘટન ઉપભોક્તા પદની ઉપર ઓછું મહત્વ આપે છે.

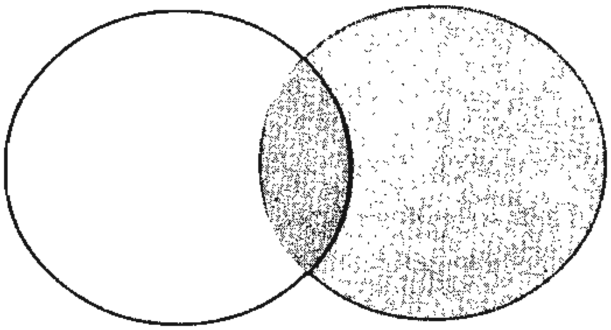
તેમ ધારણા કરવામાં આવે છે. આ ધારણાઓ સાથે જે પદનું 0.0 પ્રદાન થયેલ હોય તેની પુનઃપ્રાપ્તિ ઉપર કોઈ અસર હોવી ન જોઈએ. આ રીતે

'A1 OR B0' પદ તરીકે ધરાવતા હોય તેવા પદોનો ગણ પરત આવવો જોઈએ

'A1 AND B0' પદ તરીકે ધરાવતા હોય તેવા પદોનો ગણ પરત આવશે.

'A1 NOT B0' ગણ એ પણ પરત.

આ બાબત સૂચવે છે કે પદ ' B' નો ભાર 0.0 થી 1.0 સુધી જાય છે પરિણામે સામાન્ય રીતે બુલીયન પ્રક્રિયાથી તૈયાર કરેલ સેટ તમામ પદોના ગણમાંથી જે ગણ પદ ' A' ધરાવતા હોય તેવા ગણમાં બદલાય છે. આ પ્રક્રિયા આકૃતિ 18.1 માં દર્શાવેલી વેન આકૃતિ દ્વારા દશ્યમાન કરી શકાય છે.



આકૃતિ 18.1 વેન આકૃતિ

'A1' OR'B1' નું ચોક્કસ અર્થઘટન વેન આકૃતિના તમામ વિસ્તારમાં આવેલ તમામ એકમોનો સમાવેશ કરશે. 'A1 or B0' માં જે ફક્ત 'A' માં હોય તેવા એકમો હશે (અર્થાત્ સફેદ અને શ્યામ રંગવાળા વિસ્તાર). જે 'B NOT A' માં ન હોય તે સિવાયના તમામ એકમો (રાખોડિયા રંગનો વિસ્તાર). આ રીતે શોધ પદ 'B' નું મૂલ્ય 0.0 થી 1.0 સુધી જાય છે. 'B NOT A' ના એકમો સમપ્રમાણમાં ઉમેરાતા જશે અને 1.0 એ તમામ એકમો ઉમેરાઈ જશે.

તે જ રીતે ચોક્કસ અર્થઘટનમાં 'A1 and B1' કાળા ટપકાવાળો વિસ્તાર તમામ એકમોનો સમાવેશ કરશે. ઉપર વર્ણવ્યા મુજબ 'A1 and B0' માં હોય તે એકમોનો હશે. આ રીતે શોધ પદ 'B' નું મૂલ્ય 1.0 થી 0.0 સુધી જાય છે, 'A and B' (સફેદ વિસ્તાર) ના એકમો સમપ્રમાણમાં ઉમેરાતા જશે.

અંતિમ 'A1 NOT B1' **નું ચોક્કસ અર્થઘટન વિસ્તાર છે જ્યારે** 'A1 and B0' એ નો બધો ભાગ છે. આ રીતે 'B' નું મૂલ્ય 0.0 થી 1.0 સુધી જાય છે. એકમો 'A and B' માંથી સમપ્રમાણમાં ઉમેરાતા જશે અને 1.0 એ તમામ એકમો ઉમેરાઈ ગયા છે.

ભારિત મૂલ્યમાં કયા એકમો ઉમેરવા અથવા બાકાત રાખવા તે નક્કી કરવું એ અંતિમ પ્રશ્ન છે.

18.5.2 અસ્પષ્ટ તાર્કિક નમૂનો (Fuzzy Logic Model)

માહિતી પુનઃપ્રાપ્તિ પદ્ધતિને સોફ્ટવેર અંગ હોય છે જેને માહિતી એકમોમાં હસ્તક્ષેપ કરવા લક્ષણો અને કાર્યો હોય છે. તેથી વિરુદ્ધ ડેટા સંપુટ સંચાલન પદ્ધતિ માળખાયુક્ત ડેટા હસ્તગત કરવા મહત્તમ ઉપયોગ કરે છે. અહીં માહિતીને સ્પષ્ટ પાઠ તરીકે ઓળખવામાં આવે છે. અહીં અસ્પષ્ટ પદ પાઠના એકમોના સર્જન ઉપર લઘુત્તમ માનાંકો અથવા નિયંત્રકોમાંથી પરિણામો સૂચિત કરવા ઉપયોગ થાય છે. લેખક વિભાવનાઓ વિચારો અને સારાંશોની સાથે સહાયક હકીકતો રજૂ કરે છે. આ રીતે શબ્દભંડોળ અને એકમોની પદ્ધતિમાં લઘુત્તમ પધ્ધતિ એકસૂત્રતામાં હોય છે. શોધકે આ બધા પદો, શોધ પ્રશ્નમાં સ્પષ્ટ કરવાની શક્યતાઓ સર્વજ્ઞ બનવું પડશે.

અસ્પષ્ટ તર્કશાસ્ત્ર મૂલ્યો સાચાં અને ખોટાં તથા આ બંને વચ્ચે મૂલ્યોને સહાય પૂરી પાડે છે.

અસ્પષ્ટ તર્કશાસ્ત્રની વિભાવના પ્રો. લોફ્ટી એ માદેહ દ્વારા દાખલ કરવામાં આવી હતી. અસ્પષ્ટ તર્કશાસ્ત્રનો મૂળભૂત હેતુ કુદરતી ભાષા પ્રક્રિયાની નજીક હોય તેવો નમૂનો વિકસાવવાનો છે. તે અનિશ્ચિતતામાં સંદિગ્ધતાની સાથે બિનચોકસાઈ સંલગ્ન થયેલ હોય તેવા પ્રકારના નમૂનાઓ તૈયાર કરવા યોગ્ય સાધન છે.

અસ્પષ્ટ પુનઃપ્રાપ્તિ દાખલ કરેલ શોધ પદના જેવા જ શબ્દોના અક્ષરોનું સ્થાન શોધવાની ક્ષમતા ધરાવે છે. આ કાર્ય પ્રાથમિક રીતે શબ્દોના અક્ષરોમાં રહેલી ક્ષતિઓ સરખી કરવા ઉપયોગમાં લેવામાં આવે છે. અસ્પષ્ટ પુનઃપ્રાપ્તિ ઘટતી જતી ચોકસાઈને જોખમે પુનઃપ્રાપ્તિ આંક વધારે છે. અસ્પષ્ટ પુનઃપ્રાપ્તિ પ્રક્રિયામાં શોધપદ વિસ્તરણમાં શબ્દ લંબાઈ સરખી હોય અને અક્ષરોના સ્થાન દાખલ કરેલ પદના જેવા જ હોય અને ડેટા સંપુટમાં હોય તેવા પદોને વધુ ભાર આપીને અન્ય પદોનો સમાવેશ કરે છે. 'Computer' શબ્દની અસ્પષ્ટ શોધ સ્વયં રીતે માહિતી ડેટા સંપુટોમાંથી, 'Computer' 'Compiter' 'Computter' 'Compute' શબ્દોનો સમાવેશ કરશે. પુરક રીતે વધારામાં સૂચિત વૈકલ્પિક સ્પેલિંગ શોધશે અને તે બીજા અર્થમાં માન્ય હશે તો શોધમાં ઓછા ક્રમે સમાવેશ કરશે અથવા સદંતર સમાવેશ કરશે નહિ (દા.ત. Commuter). પધ્ધતિ શોધ પ્રશ્નના વિસ્તરણમાં સમાવેશ થતા થોડાક પદોનું મહત્તમ સંખ્યામાં સ્પષ્ટીકરણ કરવા અનુમતિ આપે છે.

અસ્પષ્ટ પુનઃપ્રાપ્તિની પધ્ધતિમાં મહત્તમ ઉપયોગિતા તે જે અક્ષરો પ્રકાશ દ્વારા ઓળખાયેલા છે. તેને સ્વીકારે તેમાં હોય છે. પ્રકાશ દ્વારા અક્ષરો (લક્ષણો) ઓળખવા (OCR) ની પ્રક્રિયામાં પ્રણાલિકાગત પ્રત (hard copy) નું દ્વિઅંકી પધ્ધતિ (Binary) છાપમાં સ્કેન કરવામાં આવે છે (OCR પેટર્ન ઓળખવાની પ્રક્રિયા છે. જે સ્કેન કરેલ છે તેને અર્થપૂર્ણ ઉપક્ષેત્રમાં અલગ પડે છે. ઘણીવાર અલગ પાડેલ ક્ષેત્રનો વિસ્તાર તે સ્વીકારીને એક લક્ષણ નક્કી કરે છે. (OCR પદ્ધતિ લક્ષણો નક્કી કરશે અને કમ્પ્યુટરના આંતરિક સંકેતોમાં ભાષાતંર કરશે. પ્રણાલિકાગતપ્રતની અસલ ગુણવત્તા આધારિત આ પ્રક્રિયા ભૂલ ઓળખ લક્ષણો દાખલ કરે છે. નિવેશની સારી ગુણવત્તા પુનઃપ્રાપ્તિની 90 થી 99 ટકા સુધીની ચોકસાઈ પ્રાપ્ત કરે છે. જો કે આ સમગ્ર અષ્પષ્ટ પુનઃપ્રાપ્તિમાં લક્ષણો (અક્ષરો) ની ભૂલો છે જે સરવાળા એકમોના અસંખ્ય લક્ષણો (અક્ષરો) નું સ્થાન સરભર કરવા અનુમતિ આપે છે.

18.5.3 ગણ સિદ્ધાંત આધારિત મોડેલ (Set Theoritic Model)

માહિતી પુનઃપ્રાપ્તિનો ગણ સૈદ્ધાંતિક અભિગમ માહિતી વિનંતીઓ ઉપર આધારિત છે જેની રચના સામાન્ય રીતે સંગ્રહની પસંદગી દ્વારા અથવા ઓળખ એકમોના ગણ અથવા ચાવીરૂપ શબ્દો દ્વારા થાય છે. ચાવીરૂપ શબ્દોનો ગણ એક પછી એક રીતે સંગ્રહિત નોંધોમાંથી પસંદગી કરેલ પુનઃપ્રાપ્તિના ઉપગણ તરફ લઈ જાય છે. પુનઃપ્રાપ્તિ સિદ્ધાંતના મૂળભૂત ડેટા આ અભિગમમાં એકમોના વર્ણનના ગણ અને તેના અનુવર્તીક નોંધો વચ્ચે અસ્તિત્વ ધરાવતા સંબંધ દ્વારા પૂરા પાડવામાં આવે છે.

18.5.4 સદિશ અવકાશ મોડેલ (Vector Space Mode) (VSM)

ઘણીવાર ભારિત પદ્ધતિઓની સદિશયુક્ત માહિતી પુનઃપ્રાપ્તિ પદ્ધતિઓ તરીકે ચર્ચા કરાય છે. આ સંગઠન (જોડાણ) કોર્નોલ યુનિવર્સીટીમાં ડો. ગેરાલ્ડ સાલ્ટોન

દ્વારા Smart પધ્ધતિના થયેલ સર્જનમાંથી આવે છે. આ પધ્ધતિ માહિતીની ઓળખ માટેના પાયા તરીકે ભારને મહત્વ આપે છે અને આ ભારને સદિશ સ્વરૂપમાં સંગ્રહ કરે છે. સદિશ નમૂના આધારિત પદ્ધતિઓમાં, દરેક એકમનો અર્થ સદિશ તરીકે પ્રદર્શિત કરવામાં આવે છે. સદિશ એ મૂલ્યોના ગણનું એક પરિમાણ છે જ્યાં ગણમાં દરેક મૂલ્યનો ક્રમ/સ્થાન નક્કી કરવામાં આવે છે અને કોઈ ખાસ નિર્ધારિત સ્થાન દર્શાવે છે. દરેક સદિશ એક પ્રલેખ પ્રદર્શિત કરે છે અને સદિશમાં દરેક સ્થાન ડેટા સંપુટમાં જુદા જુદા અનન્ય શબ્દ (પ્રક્રિયાનો પ્રતિનિધિ) પ્રદર્શિત કરે છે સદિશમાં મૂલ્યોના નિર્ધારિત સ્થાનના બે અભિગમ છે. દ્વિઅંકી અને ભારિત. દ્વિઅંકી અભિગમમાં નિર્ધારિત સ્થાન શૂન્ય અથવા એક મૂલ્ય ધરાવે છે. એક એકમમાં પ્રક્રિયાના પ્રતિનિધિત્વ અસ્તિત્વની રજૂઆત કરે છે. ભારિત અભિગમમાં નિર્ધારિત સ્થાન, જટિલ રીતે, તમામ વાસ્તવિક ધન સંખ્યાઓનો ગણ છે. એક એકમનો અર્થ પ્રદર્શિત કરવામાં દરેક પ્રક્રિયા પ્રતિનિધિત્વનું મૂલ્ય, તે પ્રક્રિયા પ્રતિનિધિનું સાપેક્ષ મહત્વ ધરાવે છે. દરેક સ્થિતિ માટે પ્રદાન કરેલ મૂલ્ય એ પ્રલેખના પદનો ભાર છે. શૂન્ય મૂલ્ય દર્શાવે છે કે તે શબ્દ પ્રલેખમાં નથી. છેલ્લા 30 વર્ષોથી પધ્ધતિ અને તેને સંબંધિત સંશોધન પરિણામોનો વિકાસ થઈ રહ્યો છે. શોધ પ્રશ્નોનું ભાષાંતર સદિશ સ્વરૂપમાં થઈ શકે છે. શોધ પ્રક્રિયા, શોધ પ્રશ્ન સદિશ અને પ્રલેખ સદિશ વચ્ચેના અંતરની ગણતરીથી પૂર્ણ થાય છે. ભારનો ઉપયોગ પણ એકમનો ક્રમ નક્કી કરવા માટે આધાર પૂરો પાડે છે. સદિશ અભિગમ અવકાશ નમૂનાનો ઉપયોગ કરીને ગાણિતિક અને ભૌતિક રજૂઆત માટે અનુમતિ આપે છે.

ડેટા સંપુટમાં થતા ગતિશીલ પરિવર્તનો તથા ભાર આપતા પરિબળોના સામાન્ય પ્રશ્નોથી વધારામાં સદિશ નમૂનામાં એક એકમના ખાસ પ્રક્રિયા પ્રતિનિધિ માટેના ભાર પ્રદાન કરવામાં વધારાના પ્રશ્નો ઉદ્ભવે છે. કોઈ એક એકમમાં વધુ વિષયની ચર્ચા કરવામાં આવેલ હોય ત્યારે સદિશ અવકાશ નમૂનામાં મુખ્ય પ્રશ્ન ઉદ્ભવે છે. જ્યારે સદિશમાં દરેક પરિમાણ અન્ય પરિમાણથી સ્વતંત્ર હોય છે ત્યારે પદો વચ્ચેના સહસંબંધ પરિબળો જોડવાનો અન્ય વિકલ્પ મળતો નથી.

સદિશ અવકાશ નમૂનાની પ્રક્રિયા ત્રણ તબક્કામાં વિભાજિત કરી શકાય છે. પ્રથમ તબક્કો પ્રલેખનું નિર્દેશીકરણ કરવાનો છે જેમાં વિષય વસ્તુ ધારણ કરતા પદો પ્રલેખના પાઠમાંથી ઉદ્ધૃત કરવામાં આવે છે. અલબતતે સ્પષ્ટ છે કે પ્રલેખના ઘણા શબ્દો જેવા કે The, Is, Are, In, To, Of, વગેરે વિષયવસ્તુ વર્ણન કરતા નથી. તેમને બીન સાર્થક શબ્દો અથવા બંધિત શબ્દો (Stop words) તરીકે ઓળખવામાં આવે છે. સ્વયં સંચાલિત નિર્દેશીકરણની પધ્ધતિમાં આ શબ્દો પ્રલેખ સદિશમાંથી દૂર કરવામાં આવે છે. તેથી પ્રલેખની રજૂઆત માત્ર વિષયવસ્તુ ધારણ કરતા પદોથી જ થશે. સામાન્ય રીતે પ્રલેખમાં કુલ શબ્દો 40 થી 50% શબ્દો બંધિત શબ્દો હોય છે. આ શબ્દો બંધિત શબ્દો યાદીની મદદ દ્વારા દૂર કરી શકાય છે. બીજો તબક્કો ઉપભોક્તાને સુસંગત પ્રલેખોની પુનઃપ્રાપ્તિ વધારવા નિર્દેશીકૃત પદોને ભાર આપવાનો છે. VSM એ વાસ્તવમાં બુલીયન પુનઃપ્રાપ્તિનો નમૂનો છે. જેમાં 100% પુનઃપ્રાપ્તિ સરખામણીને આધારે કરવામાં આવે છે. VSM ચોક્કસ સરખામણી સિવાય શોધપ્રશ્નને સૌથી વધુ સામ્યતાને પુનઃપ્રાપ્તિ માટે અનુમતિ આપે છે. આ રીતે VSM ચાવીરૂપ શબ્દ X પ્રલેખ શ્રેણીક (A) માં સારી રીતે સમજાવી શકાય છે. જેમાં ડેટા સંપુટના ચાવીરૂપ શબ્દો (W) પંક્તિમાં અને પ્રલેખ (D) સ્તંભમાં મૂકીએ તો શ્રેણીક આ પ્રમાણે મળશે.

		D1	D2	D3	D4	....	Dn
A =	W1	A11	A12	A13	A14		A1n
	W2	A21	A22	A23	A24		A2n
	W3	A31	A32	A33	A34		A3n

W4	A41	A42	A43	A44		A4n
---	----		----	----	------	----
---	----	----	----	------	----	
Wm	Am1	Sm2	Am3	Am4		Amn

ચાલો આપણે એક 'શૈક્ષણિક માહિતી પુનઃપ્રાપ્તિ પદ્ધતિ' ઉપર માહિતી શોધ છે તેવું ઉદાહરણ લઈએ. તે ચાર ચાવીરૂપ શબ્દો W1, W2, W3, W4 નો ઉપયોગ કરે છે. ડેટા સંપુટમાં શોધ કર્યા બાદ તે છ પ્રલેખ A1, A2, A3, A4, A5, and A6 મેળવે છે. પૃથક્કરણ કર્યા બાદ તે શોધી કાઢે છે કે પ્રલેખ A1 ફક્ત W1 વિશે ચર્ચા કરે છે. લેખ A2, W2 ની એક તૃતીયાંશ અને W4 ની બે તૃતિયાંશ ચર્ચા કરે છે. લેખ A3, W1 ના 20% અને W3 માં 30% અને W4 માં 50% ભાગમાં વહેંચાયેલ છે, લેખ A4, W1 ના 100% અને W4 ના 30% ભાગમાં વહેંચાયેલ છે, લેખ A5, W2 વિશે 80% અને W3 વિશે 20% ચર્ચા કરે છે અને લેખ A6 ફક્ત W4 વિશે ચર્ચા કરે છે. હવે આ બાબત 4 × 6ના શ્રેણીકમાં દર્શાવી શકાય છે.

A =

	A1	A2	A3	A4	A5	
W1	1.00	0.00	0.20	0.60	0.00	0.00
W2	0.00	0.33	0.00	0.10	0.80	0.00
W3	0.00	0.00	0.30	0.00	0.20	0.00
W4	0.00	0.67	0.50	0.30	0.00	1.00

VSM એ પુનઃપ્રાપ્તિનો નમૂનો છે જેની પુનઃપ્રાપ્તિ સંપૂર્ણ વિશાળ વર્ગની બનેલી હોય છે. દરેક વર્ગ નિર્દેશીકરણ પધ્ધતિ અને પુનઃપ્રાપ્તિના કાર્યનો બનેલો હોય છે. નિર્દેશીકરણ પદ્ધતિ વર્ણનાત્મક સદિશનું સર્જન કરે છે અને પુનઃપ્રાપ્તિ કાર્ય શોધ પ્રશ્ન વર્ણન સદિશ સાથે પ્રલેખ વર્ણન સદિશની તુલના દ્વારા પુનઃપ્રાપ્તિ મૂલ્યનું સર્જન કરે છે. VSM નું કાલ્પનિક ચિત્ર આકૃતિ 18.2 માં આપેલ છે. માહિતી મેળવનાર અંગે ધારણા કરવામાં આવે છે કે તેને માહિતીની જરૂરિયાત છે જે શોધ પ્રશ્નની રચના કરે છે. શોધ પ્રશ્ન q અને પ્રલેખ dj નું બે તબક્કામાં નિર્દેશીકરણ થાય છે. પ્રથમ શોધ પ્રશ્ન q પ્રલેખ dj માં યોગ્ય નિર્દેશીકરણ પદોનાં ટપકાં પાડવામાં આવે છે. દ્વિતીય ક્રમે શોધપ્રશ્ન વર્ણન અને પ્રલેખ વર્ણન ભારિત નિર્દેશીકૃત લક્ષણોનો ગણ છે તેને ભાર પ્રદાન કરવામાં આવે છે. શોધપ્રશ્ન વર્ણન અને પ્રલેખ વર્ણન સરખાવવામાં આવે છે અને દરેક જોડ માટે આંકનું સર્જન કરવામાં આવે છે. આ આંકને પુનઃપ્રાપ્તિ સ્થિતિ મૂલ્ય (Retrieval status value RSVs) તરીકે ઓળખવામાં આવે છે. દરેક શોધપ્રશ્ન માટે પ્રલેખો માહિતી શોધકને આ RSVs ના ઘટતા જતા ક્રમ પ્રમાણે રજૂ કરવામાં આવે છે.

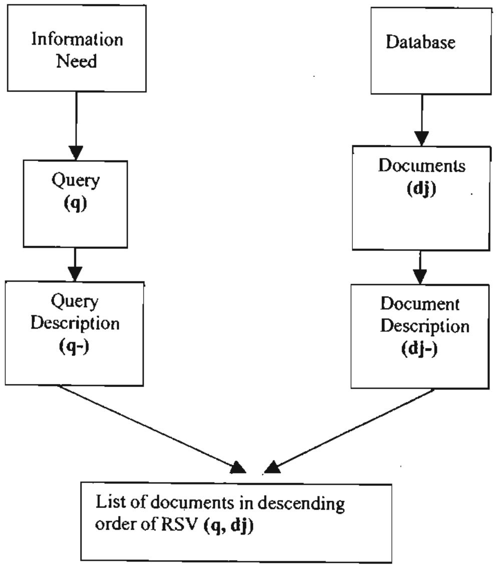


Fig. 18.2 : Vector Space Model

પ્રલેખનો અર્થ પ્રલેખની રચના કરતા પદોમાંથી નક્કી કરી શકાય છે તેવી આધાર ભૂમિકા ઉપર VSM આધારિત છે. પ્રલેખ સંગ્રહમાં ચાવીરૂપ શબ્દ પ્રલેખ સદિશની રચના કરે છે જે પ્રલેખ d માં એકવાર અથવા અનેકવાર સ્થાન પામતા પદો i ને પ્રદર્શિત કરે છે તે જ રીતે શોધ પ્રશ્ન, શોધ પ્રશ્ન સદિશ વડે પ્રદર્શિત કરવામાં આવે છે જે શોધ પ્રશ્નમાં પદો કેટલીવાર સ્થાન પામે છે. તે દર્શાવે છે. પ્રલેખ સદિશ અને શોધ પ્રશ્ન સદિશ બંને પ્રલેખપદ અવકાશમાં વસ્તુનું સ્થાન પૂરું પાડે છે (દર્શાવે છે) દરેક સદિશને લંબાઈ અને ખૂણો નિશ્ચિત બિંદુ સંદર્ભમાં એક પરિમાણ ઉપર બે સામાન્ય માપ હોય છે. બે સદિશો વચ્ચેનો ખૂણો તે બંને સદિશોના માપ અંક્ષ સંદર્ભમાં આપે છે. પ્રલેખ સદિશ જેનો ખૂણો શોધ પ્રશ્ન સદિશના ખૂણાની નજીક હોય તે શોધ પ્રશ્નની સૌથી અધિક સુસંગત પ્રલેખ પૂરા પાડતી ઉત્તમ પસંદગી છે. બે સદિશો વચ્ચેનું માપ કોટિ જ્યાં ખૂણામાં મપાય છે. જો બે સદિશો વચ્ચે કોટિજ્યા ખૂણા (cousine) નું મૂલ્ય 1 હોય ત્યારે પ્રલેખ સદિશ અને શોધ પ્રશ્ન સદિશ વચ્ચેના ખૂણાનું માપ શૂન્ય અંશ હોય છે. અર્થાત્ પ્રલેખ સદિશ અને શોધ પ્રશ્ન સદિશ બંને એક જ દિશામાં હોય છે. કોટિજ્યાનું મૂલ્ય 0 નો માપ શૂન્યનો અર્થ થાય છે કે પ્રલેખ શોધ પ્રશ્ન સદિશને વિસંગત છે. આ રીતે કોટિજ્યા માપ 1 થી નજીકનો અર્થ થાય છે કે પ્રલેખ શોધ પ્રશ્નથી નજીક સંબંધ છે.

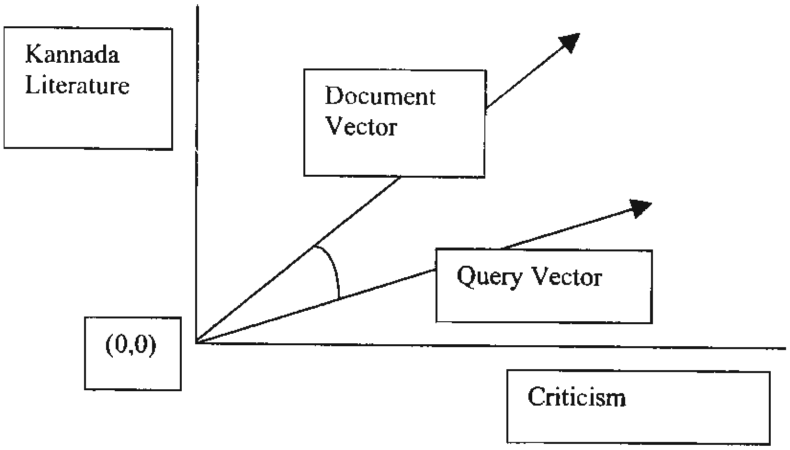


Fig. 18.3: mapping of vector spaces

જો શોધ પ્રશ્ન q ને કાલ્પનિક અવકાશમાં રેખા તરીકે સ્વીકારવામાં આવે અને પ્રલેખને પણ કાલ્પનિક અવકાશ અવકાશમાં રેખા તરીકે સ્વીકારવામાં આવે તે ભૌમિતિક રીતે થતો અંશમાં માપ દર્શાવતો બંને રેખાઓ વચ્ચેનો નક્કી થતો અંશમાં માપ દર્શાવતો ખૂણો જે પ્રલેખ શોધ પ્રશ્નના જેવા જ છે. જ્યારે ખૂણો મોટો હોય તેવા ઉદાહરણમાં પ્રલેખ શોધ પ્રશ્નના જેવા જ છે. જ્યારે ખૂણો મોટો હોય તેવા ઉદાહરણમાં પ્રલેખ શોધ પ્રશ્નથી વિસંગત છે તેમ અનુમાન કરાય છે. ખૂણાનું માપ ઘણું નાનું હોય તેવા ઉદાહરણમાં પ્રલેખ શોધ પ્રશ્નને વધુ સંગત છે તેમ અનુમાન કરાય છે. આ ઉદાહરણમાં સદિશ બિંદુ (0.0) થી કન્નડ સાહિત્ય અને વિવેચન પદો પ્રલેખ અથવા પ્રશ્નમાં કેટલીકવાર સ્થાન પામેલ છે તે નક્કી થયેલ સંખ્યાથી દોરવામાં આવેલ છે. જો 'વિવેચન' પદ એકવાર અને 'કન્નડ સાહિત્ય' પદ પ્રલેખમાં બે વાર સ્થાન પામેલ હોય તો આકૃતિ 18.3 માં સદિશ (d) પ્રાપ્ત થયેલ છે.

18.5.5 સંભાવનાત્મક પુનઃપ્રાપ્તિ નમૂનો (Probabilistic Retrieval Model)

જો કોઈ એવી ધારણા કરે કે પદ્ધતિના લક્ષણો જેવાં કે નોંધોને પ્રદાન કરેલ પદ અથવા નોંધોના સુસંગત લક્ષણો સંભાવનાત્મક સ્વરૂપમાં છે તો સંભવનાત્મક સ્વીકૃતિઓ પ્રયોજિત કરી શકાય છે અન્ય ગાણિતિક તકનિકો કે જેનો ઉપયોગ થયો હોય તેમાં નિર્ણયનો સિદ્ધાંત, માહિતી સિદ્ધાંત, આદર્શ નમૂનાનું વર્ગીકરણ (pattern classification), ગાણિતિક ભાષાશાસ્ત્ર, લક્ષણ પસંદગી પદ્ધતિઓનો સમાવેશ થાય છે. સદિશ અવકાશ નમૂના ઉપરાંત વધારામાં અન્ય પ્રભુત્વ ધરાવતા

અભિગમ સંભાવનાત્મક પુનઃપ્રાપ્તિ નમૂનાનો ઉપયોગ કરે છે. આ ક્ષેત્રમાં સૌથી વધુ સફળ થયેલ નમૂનો 'બાયેસિઅન' અભિગમ છે. આ નમૂનો માહિતીના નૈસર્ગિક સ્વરૂપવાળો તથા પુરાવા આધારિત તાર્કિક સિદ્ધાંતો ઉપર આધારિત છે. (પુરાવામાંથી નિષ્કર્ષ બહાર કાઢે છે) બાયેસિઅન અભિગમો લાંબા સમયથી માહિતી પદ્ધતિઓમાં પ્રયોજિત કરવામાં આવે છે. બાયોસિઅન અભિગમ નિર્દેશિકા પદોને ભાર આપવામાં પ્રયોજિત કરવામાં આવે છે. બાયોસિઅન અભિગમ નિર્દેશિકા પદોને ભાર આપવામાં પ્રયોજિત કરી શકાય, પરંતુ સામાન્ય રીતે પુનઃપ્રાપ્તિ પદ્ધતિના ભાગ તરીકે એક એકમ અને વિશિષ્ટ શોધ પ્રશ્ન વચ્ચેના સંબંધની ગણતરી કરવા ઉપયોગ કરવામાં આવે છે.

સંભાવનાત્મક અભિગમ, સંભાવનાના સિદ્ધાંતનો માહિતી પુનઃપ્રાપ્તિ પદ્ધતિઓમાં સીધી પ્રયોજિતતા ઉપર આધારિત છે. આનો લાભ વિકસાવેલ ઔપચારિક સંભાવના સિદ્ધાંતનો અલ્ગોરિધમ વિકાસમાં ઉપયોગ શક્ય બનાવવામાં રહેલ છે સંભાવના સિદ્ધાંતનો ઉપયોગ નૈસર્ગિક પસંદગી છે કારણ કે તાર્કિક પુરાવાનો આધાર છે.

વાન રીઝબર્ઝન (1979)ના મત પ્રમાણે માહિતી પુનઃપ્રાપ્તિનું વર્ણન અને પૃથક્કરણ કરવા માટે સંભાવનાનો સિદ્ધાંત અંતઃસ્ફૂરણ પ્રેરક છે. આ અભિગમમાં એક વ્યક્તિ પુનઃપ્રાપ્તિની સંભાવનાનો અંદાજ કાઢી શકે છે.

બીજા શબ્દોમાં પુનઃપ્રાપ્તિની સંગતતાની સંભાવના માપી શકાય છે. વાન રિઝબર્ઝને સૂચવેલ છે કે કોઈ એક ખાસ સિધ્ધાંતને લગતા પ્રલેખની સંગતતાની સંભાવવાનું માપ તે પ્રલેખ દર્શાવતા સદિશ આધારિત હોય છે. તેમણે સ્વયં સ્વિકૃત સ્વીકારેલ છે કે સુસંગત પ્રલેખોના નિર્દેશીકરણ પદોની લઢણ, વિસંગત પ્રલેખોની લઢણથી જુદી પડશે. સંભાવના સિદ્ધાંતનો અભિગમ પદ જૂથીકરણ (term clustering), આવૃત્તિ ભારાંકન (frequency weighting) , સુસંગતતા ભારાંકન (relevance weighting) અને ક્રમાંક પૃથક્કરણ કરવામાં મદદરૂપ થશે.

સંભાવનાનો સિદ્ધાંત ક્રમ આપવા તેમજ પ્રલેખને તેની સુસંગતતા હોવાની સંભાવના પ્રમાણે ક્રમ આપવા પણ ઉપયોગી પુરવાર થઈ શકે છે. રોબર્ટસને (1978) દર્શાવે છે કે પ્રલેખોનો ક્રમ, પદ મૂલ્યો તથા મહત્તમ પુનઃપ્રાપ્તિ કાર્ય ઉપર આધારિત હોઈ શકે છે. આમ છતાં સંભાવનાના બુલીયન વાતાવરણમાં કોઈપણ વ્યક્તિ પ્રલેખોને ક્રમ આપવા પ્રત્યન કરે તો કેટલીક મુશ્કેલીઓ બુલીયન તર્કશાસ્ત્રની પાયાગત સહજ બાબતોને કારણે ઊભી થાય છે. બુકસ્ટીને (1978) સૂચવ્યું હતું કે પુનઃપ્રાપ્તિ પ્રલેખોનો ક્રમ, પ્રલેખમાં બુલીયન રજૂઆત કેટલીક છે તેને આધારે આપી શકાય છે જે બાબત સાચી છે.

ડોસ્ઝકોસે (1978) સૂચવ્યું છે કે સંભાવના ઘટત્વ (Probability association) નો ઉપયોગ પદો કે જે અન્ય પદો સાથે જોડાયેલ હોય તો તે શોધવા થઈ રહ્યો છે. જોડાણ પ્રક્રિયા ડેટા સંપુટમાં પદોના સ્થાન અને આવૃત્તિ ઉપર આધારિત છે. પુનઃપ્રાપ્તિ નમૂનાનો સંભાવના અભિગમ નિર્દેશીકરણ લક્ષણોનું વિતરણ પ્રલેખની સુસંગતતા વિશે થોડુંક કહે છે તેની ધારણા ઉપર આધારિત છે આ અભિગમ પુનઃપ્રાપ્તિ નમૂનો સ્વીકારે છે જે સંભાવનાત્મક સિદ્ધાંત (Probability Ranking Principal) ના આધારે શ્રેષ્ઠતાનુંરૂપ અસરકારક પુનઃપ્રાપ્તિ આપે છે.

વિલિયમ કુપરે સંભાવના ક્રમ સિદ્ધાંતની આ પ્રમાણે રચના કરેલ છે. 'સંદર્ભ પુનઃપ્રાપ્તિનો પ્રતિભાવ દરેક વિનંતી ઉપભોક્તા કે જે વિનંતી દાખલ કરે છે તેની ઘટતી જતી સંભાવનાના ક્રમમાં છે જ્યાં સંભાવનાઓ અંદાજ આ હેતુ માટે પદ્ધતિ પાસે જે કઈ ડેટા ઉપલબ્ધ કરવામાં આવે તેના આધારે શક્ય હોય તેટલી વધુ ચોકસાઈથી કરે છે, ત્યારે સમગ્ર પદ્ધતિની અસરકારકતા તેના ઉપભોક્તાને સૌથી વધુ ઉત્તમ અસરકારકતા તે ડેટાની પ્રાપ્યતાને આધારે હશે.' આ રીતમાં પદ્ધતિ (માહિતી પુનઃપ્રાપ્તિ પદ્ધતિ) શોધ પ્રશ્નના જવાબની શરૂઆત પ્રલેખોની

યાદી કે જે આંકના ઘટતા જતા ક્રમમાં જે પોતે જ સંભાવના રજૂ કરે છે. અથવા સંભાવનાનું આલેખન ક્રમ જાળવી રાખે તેવા રૂપાંતરના ઉપયોગ દ્વારા પ્રદર્શિત કરે છે. આ આંક ઘણીવાર પુનઃપ્રાપ્તિ સ્થિતિ મૂલ્ય (Retrieval Status Value RSV) તરીકે ઓળખવામાં આવે છે જે નિર્દેશિકાના લક્ષણો અંગેની યોગ્ય આંકડાકીય માહિતી ધરાવતા પ્રલેખ વર્ણન ઉપર આધારિત હોય છે. આ પ્રકારનો આંક આધારભૂત પ્રાચલ કે જે વધારાના ડેટા જેવા કે થિસોરસ ઉપર પણ આધારિત હોઈ શકે છે.

સંભાવનાઓ સામાન્ય રીતે દ્વિભાગી શરત-એકમ સુસંગત છે અથવા તેનાથી ઉપર હોય છે પરંતુ માહિતી પદ્ધતિઓમાં એકમની સુંસગતતાની નિરપેક્ષ રીતે ઉપયોગીતાનું સાતત્ય જાળવવાનું કાર્ય છે. એકમોનું નિર્ગમન તેમના ક્રમ જે સંભાવના આધારિત છે તેના દ્વારા કરવામાં આવે છે પરંતુ જો ચોકસાઈ મુજબ ગણતરી કરેલ હોય છે. સંભાવનાના સિધ્ધાંતની પ્રયોજિતતામાં ઉદ્‌ભવતા પ્રશ્નોના સ્રોત, ચોકસાઈયુક્ત ડેટાનો અભાવ અને ગાણિતિક નમૂનામાં પ્રાયોજિત કરેલ ધારણાઓ છે તેમાંથી આવે છે.

18.5.6 ભાષાશાસ્ત્રીય મોડેલ (Lingustic Model)

માહિતી પુનઃપ્રાપ્તિ માટેની ભાષાકીય નમૂનામાં ભાષાઓના ગુણધર્મોના દષ્ટિબિંદુથી માહિતી પુનઃપ્રાપ્તિનો અભ્યાસ કરવામાં આવે છે. માહિતી પુનઃપ્રાપ્તિ કુદરતી તથા કૃત્રિમ ભાષાઓના લક્ષણો દ્વારા પૂરી પાડવામાં આવે છે. કુદરતી ભાષા આધારિત માહિતીના સંગ્રહની વિવિધ રીતો જરૂરી છે. માનવીય પ્રત્યાયન તે પોતે જ સંપૂર્ણ કુદરતી ભાષા છે. ટૂંકાણમાં ભાષાઓ ત્રણ પ્રકારના કાર્યો કરે છે.

(1) તે પ્રલેખો અને માહિતીના અન્ય સ્વરૂપોનાં વિષયવસ્તુ રજૂ કરે છે.

(2) ઉપભોક્તાઓની માહિતી પ્રશ્ન ભાષામાં રજૂ કરવામાં આવે છે અને

(3) ભાષાનો ઉપયોગ કમ્પ્યૂટર પ્રક્રિયામાં અને માહિતી શોધ અને પુનઃપ્રાપ્તિમાં થાય છે. ભાષા ત્રણ આધારો ઉપર કાર્યો કરે છે.

(અ) અર્થ નિર્ધારણ શાસ્ત્ર જે એક માનવીય સમૂહથી બીજાને અર્થ પહોંચાડે છે.

(બ) વાક્ય રચનાના નિયમોનો આધાર જે વ્યાકરણમાં અર્થમૂલ્યની રચના કરવામાં મદદ કરે છે.

(ક) શબ્દભંડોળ જે વાક્યો, ફકરા અને અન્ય માળખાની રચના માટે પદોના જુદા જુદા અર્થ પૂરા પાડે છે.

ભાષાનું તાર્કિક તથા વર્ગીકરણાત્મક બંધારણ શબ્દો તથા વિષયવસ્તુ વચ્ચેના સંબંધનો સંદર્ભ આપે છે. શબ્દભંડોળ સામાન્ય રીતે તાર્કિક બંધારણ સ્પષ્ટ કરે છે. આધુનિક સમયમાં શબ્દભંડોળ નિયંત્રણ, થિસોરસ નિયંત્રણ અને તફનિકી પારિભાષિક શબ્દ નિયંત્રણનો સમાવેશ કરે છે. રૂપાંતરિત વ્યાકરણ તથા વાક્યનું પૃથક્કરણ કરવાની તફનિકનો ઉપયોગ માહિતી પુનઃપ્રાપ્તિ માટે ભાષાની પ્રક્રિયા કરવાની ઝડપ પૂરી પાડે છે. આ ઉપરાંત સંકલિત નિયંત્રણ સાથેની નિર્દેશીકરણ ભાષા માહિતી પુનઃપ્રાપ્તિનો મૂળભૂત નમૂનો પૂરા પાડે છે. તર્કની શોધમાં અને શોધ રજૂઆતની રચનામાં સહચારી ગણિતશાસ્ત્રનો ઉપયોગ પુનઃપ્રાપ્તિમાં ભાષા નિયંત્રણનો બીજો પ્રકાર પૂરો પાડે છે. આ ભાષાશાસ્ત્રીય નમૂનો માહિતી પુનઃપ્રાપ્તિ માટેના જરૂરી આધારની રચના કરે છે. સામાજિકશાસ્ત્રના ક્ષેત્રમાં ભાષા સંદિગ્ધાર્થની ભૂમિકા ભજવે છે કારણ કે આ ક્ષેત્રની પરિભાષા કુદરતી વિજ્ઞાન ક્ષેત્ર જેટલી સખત નિયમબદ્ધ એકસૂત્રતા નથી.

18.5.7 ગાણિતિક મોડેલ (Mathematical Model)

ગાણિતિક નમૂનો સામાન્ય રીતે કાળજીપૂર્વક પ્રશ્નનું ઔપચારિક પૃથક્કરણ કરવાની ધારણાઓની સ્પષ્ટતાઓ કરવાની પૂર્વધારણા કરે છે અને નમૂનો જે રીતે ધારણાઓ ઉપર આધાર રાખે છે તેની સ્પષ્ટતા કરે છે.

ગાણિતિક નમૂના આવશ્યક રીતે નિરૂપક ગણિતશાસ્ત્ર તથા સહચારી જોડાણો ઉપર આધારિત છે. ખાસ કરીને સ્વયં સંચાલિત નિર્દેશીકરણ અને સારકરણમાં ગુચ્છિત પૃથક્કરણ અને ગુચ્છીકરણની તક્નિકોનો ઉપયોગ પ્રાયોગિક ધોરણે થાય છે. ગણ સિદ્ધાંત અને બુલીયન તર્કશાસ્ત્રનો ઉપયોગ માહિતીના ગાણિતિક નમૂના અને ગુચ્છિકરણના સંયોગીકરણનું લક્ષણ માહિતીની રજૂઆત માટે અર્થપૂર્ણ માળખું પૂરું પાડવા પ્રયત્ન કરે છે. સાંપ્રત સમયમાં ગુચ્છ પૃથક્કરણ, આંકડાશાસ્ત્રીય કાર્યક્રમ અથવા ગુચ્છિકરણ સોફટવેરનો સમાવેશ કરે છે.

18.5.8 મનોવૈજ્ઞાનિક મોડેલ (Psychological Model)

માહિતી પુનઃપ્રાપ્તિનો મનોભાષાશાસ્ત્રીય અભિગમ માનવ મગજમાં કલ્પનાની રચના જેમાં માનવ વિચાર પ્રક્રિયા જે રીતે વિચારોને ગોઠવે છે. તપાસ સમયે તેની રજૂઆત અને તેની માગણીઓ માટે પુનઃપ્રાપ્તિ ઈજનના પ્રકારના અભ્યાસ તરફ લઈ જાય છે જ્યારે શોધ પ્રક્રિયા જ્ઞાનયુક્ત સંશોધન સાથે કમ્પ્યૂટર પ્રત્યાયન પ્રક્રિયાઓ તરફ લઈ જાય છે. બેલ્કીન, બ્રુક્સ અને ઓડી (1979) ના જ્ઞાનની અનિયમિત સ્થિતિ (Anamolous state of knowladge) ના અભ્યાસ માહિતી પુનઃપ્રાપ્તિ પ્રક્રિયાના સંબંધમાં રસદાયક સૂક્ષ્મદષ્ટિ પૂરી પાડે છે. વધુમાં સાંપ્રત સમયમાં માહિતી પુનઃપ્રાપ્તિ અને કૃત્રિમ બુદ્ધિના અભ્યાસોએ માહિતી પુનઃપ્રાપ્તિમાં સુસંગત જોડાણનો મનોવૈજ્ઞાનિક સિદ્ધાંત દાખલ કરવામાં પૂરતો પ્રકાશ પાડ્યો છે.

18.5.9 આર્થિક મોડેલ (Economic Model)

માહિતી પુનઃપ્રાપ્તિનો નમૂનો માહિતી પુનઃપ્રાપ્તિની પડતર કિંમતની અસરકારકતા અને પડતર કિંમતની કાર્યક્ષમતાના માપની આસપાસ કેન્દ્રિત છે. આ બંને માપદંડ માહિતી પુનઃપ્રાપ્તિ નિવેશ તથા સફળતાપૂર્વકના નિર્ગમનના સંબંધમાં અમલીકરણ ઉપર આધારિત છે. અનેક શોધ પ્રાપ્તિ બિંદુઓની વિચારધારાની જોગવાઈનો થઈ રહેલ ઉપયોગ માહિતી રૂપાંતરના માપ માટે તક આપે છે. માહિતી પુનઃપ્રાપ્તિનું ક્ષેત્ર જેને માહિતી માપણી માટેની બીબ્લીઓમેટ્રીક્સ અને સેન્ટોમેટ્રીક્સના અભ્યાસમાં ઉપયોગ માટે આંકડાશાસ્ત્રીય અને ગણિતશાસ્ત્રીય તક્નિકો આધારિત કેટલાક નમૂના વિકસાવ્યા છે જે આર્થિક લાભોના સહસંબંધની તક પૂરી પાડે છે તેમ છતાં માહિતી પુનઃ પ્રાપ્તિઓમાં વિવિધ અસ્પર્શીય (અદશ્ય) તત્વો જે ઓળખી શકાય તેમ નથી તેમના કારણે આર્થિક નમૂનો માહિતી પુનઃ પ્રાપ્તિનો સમગ્ર તયા અભિગમ પૂરો પાડતો નથી.

સૈદ્ધાંતિક રીતે માહિતી પુનઃપ્રાપ્તિનું નમૂનાકરણ નિર્ગમન જેવા કે ડેટા તરફથી વિવિધ વિદ્યાશાખાઓમાંની પદ્ધતિઓના અભિગમ તરફથી જોઈ શકાય છે પરંતુ વ્યવહારમાં આપણે એક પણ નમૂનો એ રીતભાત મુજબ કાર્યકરનો હોય તેવો શોધી શકતા નથી. તે હેતુ માટે જુદા જુદા સ્તરના સામૂહિક નમૂનાકરણ સમગ્ર રીતે જોઈ શકાય દા.ત. માહિતી પુનઃ પ્રાપ્તિનો ERIC (Educational Resources Information Centre) નમૂનો અગત્યના માહિતી પુનઃ પ્રાપ્તિ વિસ્તારોમાંનો એક છે જે માહિતી પુનઃ પ્રાપ્તિના તમામ અભિગમોનું સંયોગીકરણ છે (Henry and Diode, 1991). આ સૈદ્ધાંતિક નમૂના પાછળના વિચાર, માહિતી પુનઃ પ્રાપ્તિનો વિવિધ બાબતોમાં દાખલ કરવા પૃથક્કરણીય અભ્યાસમાં મદદ કરવા માટેનો છે. માહિતી પુનઃપ્રાપ્તિમાં આ વિવિધ વિદ્યાશાખાલક્ષી અભિગમ સારો આધાર પૂરો પાડે છે.

18.5.10 ઉચ્ચકક્ષીય પાઠ જોડાણ નમૂનો (Hypertext Linkage Model)

ઉચ્ચીકક્ષીય પાઠ જોડાણો વધારાના માહિતી પુનઃ પ્રાપ્તિ પરિમાણનું સર્જન કરે છે. પ્રણાલિકાગત એકમો દ્વિપરિમાણીય રચનાઓ જોઈ શકાય છે. એકમોનો પાઠ એકમોમાં માહિતી રજૂ કરવું એક પરિમાણ છે. આસપાસ ગોઠવાયેલ સંદર્ભો બીજુ તાર્કિક પરિમાણ છે. જેનો શોધ તક્નિકોમાં લઘુત્તમ ઉપયોગ હોય છે.

સાઈટેરાનો મુખ્ય ઉપયોગ એકમોની અંદર અને જૂથીકરણ એકમોમાં, વિચારધારા નક્કી કરવા પ્રયત્ન કરી રહ્યો છે. ઉચ્ચકક્ષીય પાઠ તેના વધારાના વિજાણુ એકમો સાથેના જોડાણોને એકમો કે જે વિષયવસ્તુમાં વધારો કરે છે તેમની વચ્ચેના નેટવર્કીંગ તરીકે જોઈ શકાય છે. આસપાસ જોડાણોને ગોઠવવાની ક્રિયા ઉપભોક્તાને વધારાની માહિતી માટે તુર્ત જ જોડાણવાળા એકમ ઉપર જવાની અનુમતિ આપે છે. આ વધારાના પરિમાણનો સુસંગત માહિતીનું સ્થાન દર્શાવવા કેવી રીતે ઉપયોગ કરવો તે પ્રશ્ન છે.

સાંપ્રત સમયમાં ઈન્ટરનેટ તરફ જોતાં માહિતી શોધમાં મદદ કરવા ત્રણ ક્રિયાઓ વિધિઓ છે : હસ્તગત રીતે તૈયાર કરેલ નિર્દેશિકાઓ અથવા માહિતી પુસ્તિકાઓ - ડિરેક્ટરીઝ, સ્વયં સંચાલિત તૈયાર થયેલ નિર્દેશિકાઓ અને વેબ ક્રાઉલર્સ (કૃત્રિમ બુદ્ધિ પ્રતિનિધિ) યાહૂ (http://www.yahoo.com) એ પ્રથમ પ્રકારનું ઉદાહરણ છે જેમાં માહિતી સ્રોતો (વેબનું મુખ્ય પૃષ્ઠ) નું ઉચ્ચસ્તરીય જોડાણ વંશાવલી મુજબ હસ્તગત રીતે નિર્દેશીકરણ કરેલ હોય છે. ઉપભોક્તા કોઈ એક ખાસ વિષય એકમ ઉપર વધુ ઊંડાણ માટે ઉપએકમો જેવા વંશાવલિઓ દ્વારા માહિતી સંચાલન (navigate) કરી શકે છે. કેટલાક બિંદુએ ઉપભોક્તા અંતિમ એકમો જોવાની શરૂઆત કરે છે. Lycos (http://www,lycos.com) and Alta Vista (http://www.altavista.com) આપમેળે ઈન્ટરનેટના બીજા સ્થળે (સાઈટો) જાય છે અને તે સ્થળે (સાઈટો) ઉપર સ્વયં સંચાલિત નિર્દેશીકરણ માટે પાઠ પાછો લાવે છે. Lycos સ્વયં સંચાલિત નિર્દેશીકરણ માટે દરેક સ્થળમાંથી મુખ્ય પૃષ્ઠ પરત લાવે છે. જ્યારે Alta Vista તે સ્થળે તમામ પાઠોની નિર્દેશિકા તૈયાર કરે છે.

વેબ ક્રાઉલર્સ (Webcrawler, Open text, Pathfinder) અને કૃત્રિમ બુદ્ધિ પ્રતિનિધિઓ (Coriolis Group's Netseekers) એ એવા સાધનો છે કે ઉપભોક્તાને તેના રસના વિષયને વ્યાખ્યાયિત કરવા દે છે અને તે આપમેળે ઈચ્છિત માહિતી શોધ માટે ઈન્ટરનેટના વિવિધ સ્થળોએ જાય છે. એકસૂત્રીય સ્રોત શોધ (Uniform Resources Locator URL) ઉચ્ચકક્ષીય પાઠના જોડાણનું પાઠના જોડાણનું આલેખન બીજા એક એકમમાં અથવા તે એકમમાં વિશિષ્ટ સ્થાન ઉપર કરી શકાય છે.

◆ **તમારી પ્રગતિ ચકાસો.**

4. સિદ્ધાંતો અને સાધનો આધારિત માહિતી પુનઃ પ્રાપ્તિના જુદા જુદા નમૂનાઓ લખો.

નોંધ : 1. તમારો જવાબ નીચે આપેલ જગ્યામાં લખો.

2. તમારો જવાબ આ એકમને અંતે આપેલ જવાબો સાથે મેળવો.

....

....

....

....

....

....

18.6 સારાંશ (SUMMARY)

માહિતી પુનઃપ્રાપ્તિ પધ્ધતિએ એક એવી પધ્ધતિ છે જે માહિતીનો સંગ્રહ, પુનઃપ્રાપ્તિ અને નિભાવ કરવા માટે સક્ષમ છે. માહિતી આ સંદર્ભમાં પાઠ (સંખ્યા અને ડેટા) છાપ, દશ્ય, શ્રાવ્ય અને અન્ય બહુમાધ્યમી વસ્તુઓની સંયુક્ત રચના છે. તે સોફ્ટવેર કાર્યક્રમની બનેલી હોય છે જે ઉપભોક્તાને જરૂરિયાતો માહિતી શોધવા સુવિધા પૂરી પાડે છે. માહિતી પુનઃપ્રાપ્તિ પદ્ધતિ ડેટા સંપુટ, માહિતી ડેટા સંપુટ અને માળખાગત શોધ ભાષાનો સમાવેશ કરે છે. ડેટા સંપુટ નોંધો અથવા ફાઈલોનો

સમૂહનો બનેલ છે. માહિતી સંપુટ, ડેટા સંપુટનો બનેલ છે. માળખાગત શોધ ભાષા, ડેટા સંપુટમાંથી માહિતી મેળવવા માટે તથા ડેટા સંપુટ ઉત્પાદનો જેવા કે MS-Access, SQL Server and Orcale support SQL , પ્રમાણિત ભાષા સાથેના માલિકીપણાના વિસ્તરણો છે. શોધ પ્રશ્ન નિયંત્રક ભાષાનું સ્વરૂપ લે છે તે ઉપભોક્તાને પસંદ કરવા, દાખલ કરવા, સુધારવા ડેટાનું સ્થાન શોધવા અને તેવા બીજા માટે અનુમતિ આપે છે. માહિતી પુનઃપ્રાપ્તિ પદ્ધતિઓ ઉપભોક્તાના શોધ પ્રશ્નોની પ્રક્રિયા કરે છે તે જ રીતે SQL નો ઉપયોગ કરી ઓનલાઈન ડેટા સંપુટોમાં હસ્તક્ષેપ કરે છે. આ રીતે તે અભ્યાસનું અગત્યનું ક્ષેત્ર છે. પુનઃપ્રાપ્તિ નમૂનો નિવેશ અને નિર્ગમન ઉપર આધારિત છે. ડેટા પુનઃપ્રાપ્તિ નમૂનો ઉકેલ માટે ડેટા હાથ ધરે છે અને પ્રક્રિયા કર્યા સિવાયની માહિતી અથવા માહિતીને પ્રાથમિક તબક્કા તરીકે લઈ શકાય છે. વસ્તી ગણતરી પદ્ધતિએ ડેટા પુનઃપ્રાપ્તિ પદ્ધતિ છે. માહિતી પુનઃપ્રાપ્તિ નમૂનો એક હેતુ માટે ડેટાલક્ષી છે તે કેટલાક ડેટાને એક સંબંધાત્મક માળખામાં જોડે છે. જ્ઞાન પુનઃપ્રાપ્તિ નમૂનો નિર્ણય પ્રક્રિયા અને પ્રશ્નોના ઉકેલમાં સહાય કરવા માટે કેટલાક પ્રકારની માહિતી આત્મસાત્ કરે છે. તે માહિતી પ્રક્રિયા અને સંગઠન કરવા માટેનો સુઘડ નમૂનો છે. માહિતી પુનઃપ્રાપ્તિ નમૂનાઓ સિદ્ધાંતો અને સાધનો ઉપર આધારિત છે અને અસરારક માહિતી પુનઃપ્રાપ્તિ પધ્ધતિઓ વિકસાવવા પ્રયત્ન કરે છે.

18.7 તમારી પ્રગતિ ચકાસોના જવાબો (ANSWERS TO SELF CHECK EXERCISE)

(1) માહિતી પુનઃપ્રાપ્તિ એ સંગ્રહમાંથી માહિતી પંસદ કરવાની પ્રક્રિયા છે.

(2) માહિતી પુનઃપ્રાપ્તિ તક્નિકો વિશાળ ફલક ઉપર ત્રણ કક્ષાઓમાં વિભાજીત કરવામાં આવે છે.

(અ) ચોક્કસ મેળવણીયુક્ત પુનઃપ્રાપ્તિ

(બ) ઉત્તમ મેળવણીયુક્ત પુનઃપ્રાપ્તિ

(ક) અંશતઃ મેળવણીયુક્ત પુનઃપ્રાપ્તિ

(3) માહિતી પુનઃ પ્રાપ્તિના મૂળભૂત ત્રણ નમૂનાઓ છે.

(1) ડેટા પુનઃપ્રાપ્તિ નમૂનો તે ઉકેલ માટે હાથ ધરે છે આવશ્યક છે અને પ્રક્રિયા કર્યા સિવાયની માહિતી અથવા માહિતીનો પ્રાથમિક તબક્કો લઈ શકાય છે. વસ્તી ગણતરી પધ્ધતિ એ ડેટા પુનઃપ્રાપ્તિ પદ્ધતિ છે.

(2) માહિતી પુનઃપ્રાપ્તિ નમૂનો - તે હેતુ માટે હેતુલક્ષી છે. તે કેટલાક ડેટાને એક સંબંધાત્મક માળખામાં જોડે છે.

(3) જ્ઞાન પુનઃપ્રાપ્તિ નમૂનો - તે નિર્ણય પ્રક્રિયા અને પ્રશ્નોના ઉકેલમાં સહાય કરવા માટે કેટલાક પ્રકારની માહિતી આત્મસાત્ કરે છે. તે માહિતી પ્રક્રિયા અને સંગઠન કરવા માટેનો સુઘડ નમૂનો છે.

(4) સિદ્ધાંતો અને સાધનો આધારિત માહિતી પુનઃપ્રાપ્તિના નમૂનાઓ આ પ્રમાણે છે.

- બુલીયન પુનઃપ્રાપ્તિ નમૂનો (પ્રમાણિત અને ભારિત)
- અસ્પષ્ટ તાર્કિક નમૂનો
- ગણ સૈદ્ધાંતિક નમૂનો
- સદિશ અવકાશ નમૂનો
- સંભાવનાત્મક પુનઃપ્રાપ્તિ નમૂનો
- ભાષાકીય નમૂનો
- ગાણિતિક નમૂનો
- મનોવૈજ્ઞાનિક નમૂનો
- આર્થિક નમૂનો
- ઉચ્ચકક્ષીય પાઠ નમૂનો

18.8 ચાવીરૂપ શબ્દો (KEY WORDS)

જ્ઞાનની અનિયમિત સ્થિતિ : માહિતી શોધકના જ્ઞાન આધારમાં ખાસ હકીકતોનો અભાવ, તે ખૂટતી માહિતીનો વધુ વિશાળ વિસ્તાર રજૂ કરે છે અથવા જ્ઞાનના માળખાની ઉણપ રજૂ કરે છે.

અસ્પષ્ટ પુનઃપ્રાપ્તિ : દાખલ કરેલ શોધ પદ જેવાં જ શબ્દોના સ્પેલીંગ (અક્ષરો) હોય તેવા શબ્દોનું સ્થાન શોધવાની ક્ષમતા, અસ્પષ્ટ પુનઃપ્રાપ્તિ,પુનઃપ્રાપ્તિ, રીકોલ વધારે છે. જ્યારે ચોક્કસાઈ ઘટાડે છે. શોધ પદ વિસ્તરણમાં અસ્પષ્ટ પુનઃપ્રાપ્તિ એક જ સરખો સ્પેલીંગ હોય તેવા પદનો સમાવેશ કરે છે આ કાર્યનો પ્રાથમિક રીતે શબ્દનો સ્પેલીંગ કરવામાં આવે છે.

પુનઃપ્રાપ્તિ સ્થિતિ મૂલ્યો : પુનઃપ્રાપ્તિ પ્રક્રિયામાં શોધ પ્રશ્ન વર્ણન અને પ્રલેખ વર્ણન સરખાવવામાં આવે છે અને દરેક પ્રલેખ જોડ માટે એક આંકનું સર્જન કરવામાં આવે છે આ અંકને પુનઃપ્રાપ્તિ સ્થિતિ મૂલ્યો તરીકે ઓળખવામાં આવે છે.

ભાર આપવો : ભાર આપવો એ નિર્દેશિકાના જે પદો એકમમાં વપરાયા હોય તેને મહત્વ પ્રદાન કરવાની પ્રક્રિયા છે. ભાર એ એકમમાં પ્રદર્શિત થયેલ હોય તેવી વિચારધારા નિર્દેશિકા પદો સાથે સંલગ્ન હોય તેવા પદોની માત્રા રજૂ કરે છે. ભારિત નિર્દેશીકરણ પધ્ધતિમાં નિર્દેશિકા પદોનું પ્રલેખની વિચારધારા સાથેના જોડાણની રજૂઆતને મૂલ્ય આપવાનો પ્રયત્ન કરવામાં આવે છે.

18.9 સંદર્ભો અને વિશેષ વાચન (REFERENCE AND FURTHER READING)

Belkin, N. J. Brooks, H. M. and Oddy, R. N. (1979). *Representing and classifying Analogous States of Knowledge. In : The analysis of meaning : informatics, edited* by M. Maccafferty and K. Gray. London : Aslib. Bookstein, Abraham. (1978). On *the perils of merging Boolean and weighted retrieval system. Journal of American Society for Information Science,* 29 (3), 156-8.

Brookes, Bertram C. and Grifiths, Joseph (1978). *Frequency rank distribution Journal of the American Society for Information Science,* 27 (1), 13-17. Carter, M. B. (1986). A methodology for the economic appraisal of management information. International Journal of Information Management, 193-203.

Doszkoc, Tamas E. (1978). AID : anAssociative Interactive Dicitonary of online searching. Online Review, 2 (2), 163-73.

Gopinath, M. A. (1999). Information retrieval. In : MLIS-03 course materials.Unit 13, New Delhi : Indira Gandhi National Open University.

Gopinath, M. A. (1999). ISAR Systems : Operations and design. In : MLIS-03

course materials. New Delhi : Indira Gandhi National Open University.

"Henry, G and Diodato, V. (1991). The rates of assignment of thesaurus terms in the ERIC information retrival system : an analysis of hierarchies and levels. Journal of Documentaiton, 47 (3), 276-283.

Karen, Spark Jones (1973). Linguistics and information science. New York : Academic.

Kemp, Alister (1988). Knowledge base retrieval system. London : Aslib.

Levitan, K. B. (1982). Information resources as 'Goods' in the life cycle of ʳ I information production/Journal of American Society of Information Science,33,44-54

Mock, T. L. and Vasarhelyi, M. A. (1980). A synthesis of the information economics and lens model. Journal of Accounting Research, 477-505

Murthy, S.G.K. and Biswas, R. N. (2004).

Gopinath /M.A(1999) objectives and information retrieval suffers In:MLIS-03 Course materials. New Delhi.Indira Gandhi National open Univer City

A fuzzy logic based search tecninque for digital libraries. DESIDOC Bulletin of Information Techonolgy, 24 (6), 3-10

Rjjsbergen, Van C. J. (1979). Probabilistic retrieval. In : Information retrieval, edited by C. J. Van Rijsbergen, 2nd Ed. London : Butterworths.

Robertson, Stephen E., (1978). On the nature of fuzz : a diatribe. Journal of the American Society for Information Science, 29 (6), 304-7

Salton, Genrard. (1979). Mathematics and information retrieval. Journal of Documentation, 35 (1), 1-29

Salton, G. and McGill, M. J. (1983). Introduction to modern information retrieval, New York : Me Graw-Hill.

Smith, L. C. (1976). Artifical Intelligence in Information Retrieval System. Information Processing and Management, 12 (3) 189-222.

Wikipedia. org. (2005). Information retrial, <http://en.wikipedia.org/wiki/ (information retrieval

Mcgill, Michael j. (1978). knowledge and information spaces: Inpicatins for retrieval systems internal for the American Society for information science 27(4), 205-10