THE ROLE OF INFORMATION RETRIEVAL IN KNOWLEDGE MANAGEMENT

DR. K V S N JAWAHAR BABU*; V. HARSHA VARDHAN**; J.S. ANANDA KUMAR***

*Principal,
KMM College,
Tirupati.
**Assistant Professor,
Department of MCA,
KMMIPS, Tirupati.
***Assistant Professor,
Department of MCA,
KMMIPS, Tirupati.

ABSTRACT

Information Retrieval is the art and science of retrieving from a collection of items that serves the user purpose. It is used as a field in the areas of text mining. Information Retrieval discusses ways in which data or information can be retrieved along with types of information, the models used for data retrieval and the ways to measure accuracy. Knowledge management plays a curtail role in making decisions in any organizations. To perform knowledge management information should be retrieved effectively.

KEYWORDS: Database, Information Retrieval, Knowledge Management, Structured Data.

I. INTRODUCTION

Information retrieval is the art and science of searching for information in documents, searching for documents themselves, searching for metadata which describes documents, or searching within databases, whether relational stand alone databases or hypertext networked databases such as the Internet or intranets, for text, sound, images or data. It is the art and science of retrieving from a collection of items that serves the user purpose. The main purpose is to retrieve what is useful while leaving behind what is not.

Now the world has changed, and hundreds of millions of people engage in information retrieval every day when they use a web search engine or search their email. Information retrieval is fast becoming the dominant form of information access, overtaking traditional database- style searching (the sort that is going on when a clerk says to you:

II. SCOPE

IR can also cover other kinds of data and information problems beyond that specified in the core definition above. The term “unstructured data” refers to data which does not have clear, semantically overt, easy-for-a-computer structure. It is the opposite of structured data, the
canonical example of which is a relational database, of the sort companies usually use to main- tain product inventories and personnel records. In reality, almost no data are truly “unstructured”. This is definitely true of all text data if you count the latent linguistic structure of human languages. But even accepting that the intended notion of structure is overt structure, most text has structure, such as headings and paragraphs and footnotes, which is commonly repre- sented in documents by explicit markup (such as the coding underlying web Information that is to be retrieved is of different types

UNSTRUCTURED INFORMATION

Unstructured Information represents data that doesn’t have clear structure such as Text, paragraphs, images, sounds etc

STRUCTURED INFORMATION

Structured Information is the data which has structure such as table of data in a database, some times a document can also be a structured data which has Header, footer etc

III. OBJECTIVES OF IR

Traditionally, IR has concentrated on finding whole documents consisting of written text; much IR research focuses more specifically on text retrieval – the computerized retrieval of machine-readable text without human indexing. But there are many other interesting areas:

● Speech retrieval, which deals with speech, often transcribed manually or (with errors) by automated speech recognition (ASR).

● Cross-language retrieval, which uses a query in one language (say English) and finds documents in other languages (say Chinese and Russian).

● Question-answering IR systems, which retrieve answers from a body of text. For example, the question Who won the 1997 World Series? finds a 1997 headline World Series: Marlins are champions .

● Image retrieval, which finds images on a theme or images that contain a given shape or color.

● Music retrieval, which finds a piece when the user hums a melody or enters the notes of a musical theme.

● IR dealing with any kind of other entity or object: works of art, software, courses offered at a university, people (as experts, to hire, for a date), products of any kind.

Text, speech, and images, printed or digital, carry information, hence information retrieval. Not so for other kinds of objects, such as hardware items in a store. Yet IR methods apply to retrieving books or people or hardware items, and this article deals with IR broadly, using "document" as stand-in for any type of object.
UTILITY AND RELEVANCE

A document's utility depends on three things, topical relevance, pertinence, and novelty.

A document is topically relevant for a topic, question, or task if it contains information that either directly answers the question or can be used, possibly in combination with other information, to derive an answer or perform the task. It is pertinent with respect to a user with a given purpose if, in addition, it gives just the information needed; is compatible with the user’s background and cognitive style so he can apply the information gained; and is authoritative. It is novel if it adds to the user’s knowledge. Analogously, a soccer player is topically relevant for a team if her abilities and playing style fit the team strategy, pertinent if she is compatible with the coach, and novel if the team is missing a player in her position.

Utility might be measured in monetary terms: “How much is it worth to the user to have found this document?” “How much is this player worth to us?” “How much did we save by finding this software?” In the literature, the term “relevance” is used imprecisely; it can mean utility or topical relevance or pertinence. Many IR systems focus on finding topically relevant documents, leaving further selection to the user.

Relevance is a matter of degree; some documents are highly relevant and indispensable for the user’s tasks; others contribute just a little bit and could be missed without much harm (see ranked retrieval in the section on Matching).

From relevance assessments we can compute measures of retrieval performance such as

IV. STEPS IN THE IR PROCESS

An IR system prepares for retrieval by indexing documents (unless the system works directly on the document text) and formulating queries, resulting in document representations and query representations, respectively; the system then matches the representations and displays the documents found and the user selects the relevant items. These processes are closely intertwined and dependent on each other. The search process often goes through several iterations: Knowledge of the features that distinguish relevant from irrelevant documents is used to improve the query or the indexing (relevance feedback).

INDEXING: Creating Document Representations Indexing (also called cataloging, metadata assignment, or metadata extraction) is the manual or automated process of making statements about a document, lesson, person, and so on, in accordance with the conceptual schema (see Figure 4). We focus here on subject indexing – making statements about a document's subjects. Indexing can be document-oriented – the indexer captures what the document is about, or request-oriented – the indexer assesses the document’s relevance to subjects and other features of interest to users; for example, indexing the testimonies in Figure 2 with Jewish-Gentile relations, marking a document as interesting for a course, or marking a photograph as publication quality. Related to indexing is abstracting – creating a shorter text that describes what the full document is about (indicative abstract) or even includes important results (informative abstract, summary). Automatic summarization has attracted much research interest.
Automatic indexing begins with raw feature extraction, such as extracting all the words from a text, followed by refinements, such as eliminating stop words (and, it, of), stemming (pipes Y pipe), counting (using only the most frequent words), and mapping to concepts using a thesaurus (tube and pipe map to the same concept). A program can analyze sentence structures to extract phrases, such as labor camp (a Nazi camp where Jews were forced to work, often for a company; phrases can carry much meaning). For images, extractable features include color distribution or shapes. For music, extractable features include frequency of occurrence of notes or chords, rhythm, and melodies; refinements include transposition to a different key.

Raw or refined features can be used directly for retrieval. Alternatively, they can be processed further: The system can use a classifier that combines the evidence from raw or refined features to assign descriptors from a pre-established index language. To give an example from Figure 2, the classifier uses the words life and model as evidence to assign bioinformatics (a descriptor in Google’s directory). A classifier can be built by hand by treating each descriptor as a query description and building a query formulation for it as described in the next section. Or a classifier can be built automatically by using a training set, for machine learning of what features predict what descriptors. Many different words and word combinations can predict the same descriptor, making it easier for users to find all documents on a topic Assigning documents to (mutually exclusive) classes of a classification is also known as text categorization. Absent a suitable classification, the system can produce one by clustering – grouping documents that are close to each other (that is, documents that share many features).

QUERY FORMULATION: CREATING QUERY REPRESENTATIONS

Retrieval means using the available evidence to predict the degree to which a document is relevant or useful for a given user need as described in a free-form query description, also called topic description or query statement. The query description is transformed, manually or automatically, into a formal query representation (also called query formulation or query for short) that combines features that predict a document’s usefulness. The query expresses the information need in terms of the system’s conceptual schema, ready to be matched with document representations. A query can specify text words or phrases the system should look for (free-text search) or any other entity feature, such as descriptors assigned from a controlled vocabulary, an author’s organization, or the title of the journal where a document was published. A query can simply give features in an unstructured list (for example, a “bag of words”) or combine features using Boolean operators (structured query). Examples:

**BAG OF WORDS:** (pipe tube capillary plastic polyethylene production manufacture) Boolean query: (pipe OR tube OR capillary) AND (plastic OR polyethylene) AND (production OR manufacture).

The Boolean query specifies three AND ed conditions, all of which are necessary (contribute to the document score); each condition can be filled by any of the words joined by OR; one of the words is as good as two or three. If some relevant documents are known, the system can use them as a training set to build a classifier with two classes: relevant and not relevant.
Stating the information need and formulating the query often go hand-in-hand. An intermediary conducting a reference interview helps the user think about the information need and find search terms that are good predictors of usefulness. An IR system can show a subject hierarchy for browsing and finding good descriptors, or it can ask the user a series of questions and from the answers construct a query. For buying a digital camera, the system might ask the following three questions:

- What kind of pictures do you take (snapshots, stills, ...)?
- What size prints do you want to make (5x7, 8x10, ...)?
- What computer do you want to transfer images to?

Without help, users may not think of all the features to consider. The system should also suggest synonyms and narrower and broader terms from its thesaurus. Throughout the search process, users further clarify their information needs as they read titles and abstracts.

Matching the query representation with entity representations. The match uses the features specified in the query to predict document relevance. In exact match the system finds the documents that fill all the conditions of a Boolean query (it predicts relevance as 1 or 0). To enhance recall, the system can use synonym expansion (if the query asks for pipe, it finds tubes as well) and hierarchic expansion or inclusive searching (it finds capillary as well). Since relevance or usefulness is a matter of degree, many IR systems (including most Web search engines) rank the results by a score of expected relevance (ranked retrieval). Consider the query Housing conditions in Siemens labor camps. Figure 5 illustrates a simple way to compute relevance scores: Each term's contribution is a product of three weights: The query term weight (the importance of the term to the user), the term frequency (tf) (the number of occurrences of the term in the document, synonyms count also), and the rarity of the term or inverse document frequency (idf) on a logarithmic scale.

SELECTION

The user examines the results and selects relevant items. Results can be arranged in rank order (examination can stop when enough information is found); in subject groupings, perhaps created by automatic classification or clustering (similar items can be examined side by side); or by date. Displaying title + abstract with search terms highlighted is most useful (title alone is too short, the full text too long). Users may need assistance with making the connection between an item found and the task at hand.

Relevance Feedback and Interactive Retrieval. Once the user has assessed the relevance of a few items found, the query can be improved:

The system can assist the user in improving the query by showing a list of features (assigned descriptors; text words and phrases, and so on) found in many relevant items and another list from irrelevant items. Or the system can improve the query automatically by learning which features separate relevant from irrelevant items and thus are good predictors of relevance. A simple version of automatic query adjustment is this: increase the weights of features from relevant items and decrease the weights of features from irrelevant items.
V. IR MODELS

1. THE BOOLEAN MODEL

-- The simplest retrieval model

--Queries are index terms linked by AND, OR or NOT. They are converted into disjunctive normal form, where each part is a binary weighted vector corresponding the tuple 

\((k_a, k_b, k_c)\)

--So the weights for each keyword end up either 1 or 0 – there, or not there

--Disadvantage: since weights are binary, docs are either relevant or irrelevant, there’s no further ranking.

MEASURING DOCUMENT-QUERY RELEVANCE

The Boolean model is crude -- indexed document keywords actually vary in relevance to the query.

- “Making a pie is easy. It is not rocket science.”
- ‘pie’ is of low relevance in a query for a document on rockets.
- How do we measure degree of relevance?

THE VECTOR MODEL-1

- Non-binary weighting

- the relevance of index terms to a query and to documents are quantified as a graded scale of weights

\(W_{i,q} = \text{weight associated with a query and a Document index term in the system.}\)

\(W_{i,j} = \text{weight associated with a document and a document index term in the system.}\)

HOW ARE INDEX TERMS WEIGHTED?

WEIGHTING INDEX TERMS

- The weight of an index term is proportional to its frequency in a document (term frequency or tf factor), and inversely proportional to its frequency among all documents in the system (inverse doc frequency or idf factor).
A word like “report” will show up in a relatively high number of documents, so it can’t be very useful in distinguishing this document from all others. So the word’s idf factor would be high, compared to a word like “Crotaphytus” (assuming it’s not a lizard database).

2. THE VECTOR MODEL-II

Index term weights with relation to each doc, and to the query, are stored in vectors.

\[ d_j = (w_1, j, w_2, j \ldots w_t, j) \]

- the document vector describing the relevance of each index term in the system to document \( d_j \).

\[ q = (w_1, q, w_2, q \ldots w_t, q) \]

- the query vector describing the relevance of each index term in the system to the query (\( t \) is the total number of index terms in the system).

3. THE VECTOR MODEL-III

- Document-Query Relevance is measured by the correlation between a document vector and the query vector.

- All document vectors are measured against the query vector.

- The correlation is quantified as the cosine of the angle between the two vectors.

- Similarity \( \text{sim}(d_j, q) \) will be a value that ranges from 0 to 1.

- The IR system can set a threshold somewhere in between and return only the docs above that threshold of similarity.

VI. IR SYSTEM EVALUATION

IR systems are evaluated with a view to improvement (formative evaluation) or with view to selecting the best IR system for a given task (summative evaluation). IR systems can be evaluated on system characteristics and on retrieval performance. System characteristics include the following:

- The quality of the conceptual schema (Does it include all information needed for search and selection?);

- The quality of the subject access vocabulary (index language and thesaurus) (Does it include the necessary concepts? Is it well structured? Does it include all the synonyms for each concept?);
• The quality of human or automated indexing (Does it cover all aspects for which an entity is relevant at a high level of specificity, while avoiding features that do not belong?);

• The nature of the search algorithm; the assistance the system provides for information needs clarification and query formulation; and the quality of the display (Does it support selection?).

Measures for retrieval performance (recall, discrimination, precision, novelty) were discussed in the section Relevance and IR system performance. Requirements for recall and precision vary from query to query, and retrieval performance varies widely from search to search, making meaningful evaluation difficult. Standard practice evaluates systems through a number of test searches, computing for each a single measure of goodness that combines recall and precision, and then averaging over all the queries. This does not address a very important system ability: the ability to adapt to the specific recall and precision requirements of each individual query. The biggest problem in IR evaluation is to identify beforehand all relevant documents (the recall base); small test collections have been constructed for this purpose, but there is a question of how well the results apply to large-scale real-life collections. The most important evaluation efforts of this type today are TREC and TDT (see Further Reading).

**PRECISION**

Precision is the fraction of the documents retrieved that are relevant to the user's information need. \[
\text{Precision} = \frac{|\{\text{relevant documents} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}
\]

In binary classification, precision is analogous to positive predictive value. Precision takes all retrieved documents into account. It can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n.

Note that the meaning and usage of "precision" in the field of Information Retrieval differs from the definition of accuracy and precision within other branches of science and technology.

**RECALL**

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

\[
\text{Recall} = \frac{|\{\text{relevant documents} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}
\]

In binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query.
It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

**FALL-OUT**

The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available:

\[
\text{Fall-out} = \frac{|\{\text{non-relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{non-relevant documents}\}|}
\]

In binary classification, fall-out is closely related to specificity \((1 - \text{specificity})\). It can be looked at as the probability that a non-relevant document is retrieved by the query.

It is trivial to achieve fall-out of 0% by returning zero documents in response to any query.

**F-MEASURE**

Main article: F-score

The weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is:

\[
F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}
\]

This is also known as the \(F_1\) measure, because recall and precision are evenly weighted.

The general formula for non-negative real \(\beta\) is:

Two other commonly used F measures are the \(F_2\) measure, which weights recall twice as much as precision, and the \(F_{0.5}\) measure, which weights precision twice as much as recall.

**VII. KNOWLEDGE MANAGEMENT**

While describing knowledge management concept, it should be considered the process of knowledge in organizations such as storing, collecting, structuring, sharing, controlling, creating, disseminating, codifying, using and exploiting. These processes which are based on managing knowledge have been evaluated non-hierarchical order in the knowledge management literature and describe part of the knowledge management definition. Therefore, this study is aimed to deconstruct of knowledge management definition in organization and investigate how organization can apply knowledge management by taking in to consideration all details of knowledge processes in hierarchical order and how does this model can be maintained. For effective management of knowledge in organizations; the Chief Knowledge Officers should
focus on the KM Life Cycle Model by considering knowledge processes. There are four processes of knowledge management which are consist of capturing, organizing, refining and transferring. The capturing phase deals with knowledge capture and includes e-mail, audio files, digital files and the like. After the capturing phase, captured data or information should be organized in a way that can be retrieved and used to generate useful knowledge. One can use indexing, clustering, cataloguing, filtering, codifying and other methods can be used. The third process of knowledge management is refining. Data mining can be applied in this phase. Data mining takes explicit knowledge found in databases and transforms it into tacit knowledge. The last phase of knowledge management process is transfer. Knowledge should be disseminated or transferred by making knowledge available to employees via tutorials or guidelines for effective use.

Knowledge management process in 4 steps. These four KM processes are supported by a set of seven KM sub processes.

The first step is ‘knowledge discovery’ which may be defined as the development of new tacit or explicit knowledge from data and information or from the synthesis of prior knowledge. The knowledge discovery step has two subprocesses as combination and socialization. The discovery of new explicit knowledge relies most directly on combination, whereas the discovery of new tacit knowledge relies most directly on socialization.

The second step of knowledge management process is ‘knowledge capture’ which may be defined as the process of retrieving either explicit or tacit knowledge that resides within people (individuals or groups), artifacts (practices, technologies or repositories) or organizational entities (organizational units, organizations, interorganizational networks). This step’s subprocesses are externalization that involves converting tacit knowledge into explicit forms such as words, concepts, visuals, or figurative language and internalization that transforms of explicit knowledge into tacit knowledge.

‘Knowledge sharing’ is the third step of knowledge management processes. Tacit or explicit knowledge is communicated to other organizational participants in this step and three important clarifications are in order. First, knowledge sharing means effective transfer, so that the recipient can understand it well enough to act on it. Second, what is shared is knowledge instead of recommendations based on the knowledge. Third, knowledge sharing may take place across individuals as well as across groups, departments, or organizations. As a result knowledge sharing step has got two sub processes in the names of socialization and exchange. Socialization was explained in the Awed and Ghaziri’s study and exchange focuses on the sharing of explicit knowledge. It is used to communicate or transfer explicit knowledge between individuals, groups and organizations. The last step of managing knowledge management process is ‘knowledge application’. This means making Decisions and performing task perfectly in organizations. It requires knowledge utilizations benefits from two processes that do not involve the actual transfer or exchange of knowledge between the concerned individual’s routines and directions that are consist of sub processes in this step. Directions refer to the process through which individuals possessing the knowledge direct the action of another individual without transferring to that person the knowledge underlying the direction. Routines involve the utilization of knowledge embedded in procedures, rules and norms that guide future behavior. There are seven
components in this model: organizing, sharing, adapting, using, creating, defining, and collecting. The study argues that if company would like to determine what they know, it firstly must espouse this model in turn. For instance, without collecting knowledge, creating stage is not possible to survive in organization. Each of the stages depends on the other and not only exhibit in hierarchical order but also interact between each other technologies such as knowledge creation, storage/retrieval, transfer and application. But this information technology perspective is not only limited to our understanding about managing knowledge but also misunderstand the concept of knowledge management application in organizations.

VIII. A NEW LIFE CYCLE MODEL FOR KNOWLEDGE MANAGEMENT

In accordance with knowledge management literature, five basic processes can be considered by managing knowledge. These can be defined as creating, sharing, structuring, using, and auditing in turn that is called “knowledge management life cycle” model. This model makes us to understand knowledge management processes in hierarchical order. Each model is explained in the following paragraphs.
I. KNOWLEDGE CREATING The first stage of managing organizational knowledge requires entering the ‘knowledge kitchen’.

In other words, exploring knowledge creating stage where can be processed in organization leads us to focus which individual, group, and department on. Because if knowledge can not be created in organization; neither sharing nor auditing knowledge can be carried out.

There are too many knowledge creators in knowledge kitchen due to the fact that organization can not create collective knowledge by itself. Thus, organizational participants create knowledge through their intuition, ability, skills, behaviors, and work experiments. ‘Key players, departments and their interactivity can play a critical role in creating knowledge in organization’.

Two forms of knowledge can appear while creating knowledge. These are tacit and explicit knowledge which are embedding in organization’s products, services and work processes after creating. “The explicit knowledge can be defined as words, diagrams, or photographs that can not convey information that can be understood by direct pointing, or demonstrating, or feeling”. Explicit knowledge is technical or academic data or information that is described in formal language, like manuals, mathematical expressions, copyright and patents. It is gained through formal education or structured study and codifies, stores, hierarchy of databases and accesses with high quality, reliable, fast information retrieval systems. Therefore, explicit knowledge is easy to structure and retrieval.

Tacit knowledge includes relationship, norms, values, and standard operating procedures. Because tacit knowledge is much harder to detail, copy, and distribute, it can be sustainable sources of competitive advantage. What increasingly differentiates success and failure is how well you locate, leverage, and blend available explicit knowledge with internally generated tacit knowledge. Nonaka emphasizes two dimensions of tacit knowledge. These are technical and cognitive. Technical dimensions covers the kind of informal personal skills of crafts often referred to as “know-how”. ‘Knowing-how’ is characteristic of the expert, who acts, makes judgments, and so forth without explicitly reflecting on the principles or rules involved Cognitive dimension consists of beliefs, ideals, values, schemata, and mental models. Hereby, the basic goal of knowledge management is to convert from tacit to explicit form of knowledge in organization through following participant’s human and social information processing. There are some barriers that take place in knowledge creating stage. First is individual and the second is organizational level. The first barrier contains beliefs that people can not easily adapt to organization enough and the second is the need for a legitimate language, organizational stories, procedures and company’s paradigm.

2. KNOWLEDGE SHARING The second important stage of knowledge management life cycle is knowledge sharing. The tools for effective knowledge sharing as follows:

- formal social communication network,
- informal social communication network
- teamwork
• communities of practices
• organizational learning
• rumors and
• formal structured technological communication networks (e-mail, mobile communications, teleconferences, videoconferences, etc.). Knowledge sharing involves creating knowledge by individuals and groups with their interactivity and connectivity in organizations. Knowledge sharing is carried out by social and technical communication channels.

2.1. CONSTRUCTING SOCIAL COMMUNICATION INFRASTRUCTURE This infrastructure requires an effective interactivity between workers in informal ways. The main purpose of this infrastructure is not only converting tacit knowledge into explicit forms in the individual level, but also transmitting message from bottom to up and up to bottom in appropriate positions in the organizational level. Three types of network should be constructed while designing social communication channel: oral communication, written communication, and nonverbal communication. The chief means of conveying messages is oral communication. Speeches, formal one-on-one and group discussions, and the informal rumor mill or grapevine are popular forms of oral communication. Written communication includes memos, letters, electronic mail, fax transmissions, organizational periodicals, notices placed on bulletin boards, or any other device that is transmitted via written words or symbols. Nonverbal communication entails body movements, the intonations or emphasis we give the words, facial expressions, and the physical distance between the sender and receiver. Knowledge management prefers all three forms of organizational communication because the effective knowledge management system requires all forms of knowledge such as written/verbal, explicit/tacit, audio/visual in organization.

2.2. CONSTRUCTING TECHNICAL COMMUNICATION INFRASTRUCTURE Technical communication infrastructure refers to information and communication technology. Information networks, technical communities of practice, internet, web-based networks, intranets, and extranets should be considered in this context. Participants can share their expertise knowledge through e-mail, in-group computerized communication networks, databases, telephone conversations. Technical communication infrastructure which is known as formal communication networks provide in sharing, structuring, classifying and organizing explicit/tacit knowledge in the environment. The best technological infrastructure for the best knowledge management application allows knowledge flow continuously, mapping information correctly, distributes data sources equally, exchanges information timely and contains intelligence agents and network mining. Notwithstanding designing effective infrastructure in technological perspective permits to construct a good communities of practice that is “a group of practitioners who share a common interest in a specific area of competence and are willing to work together”

2.3. KNOWLEDGE STRUCTURING After constructing a perfect infrastructure system for knowledge sharing; data, information and knowledge should be structured in order to store in
organization’s database for the future needs. Structuring knowledge is based on sorting, organizing, codifying, analyzing, and reporting information that provides information retrieval what organization needs in the future. Knowledge structuring is frequently processed by technical communication infrastructure which “includes structuring databases, organizing data for analyzing, taxonomy of data, clustering/managing databases”.

Knowledge structuring categorizes data and information through certain types of classification tools and enables for retrieving this information timely. This means that mapping, storing and retrieving information are three important components of knowledge structuring. First is mapping information that refers to determine organizational information sources and what participant knows. A good knowledge mapping benefits from second hand information to the first hand and making knowledge inventory available to overall organization. Second is information storing that contains knowledge repositories such as databases, data warehouses, and information centers and indicates electronic environment of organizational memory. Third is the most critical factor in structuring knowledge that is called information retrieval. In this stage, knowledge is stored and retrieved via information retrieval systems such as surrogates, user interface, Boolean logic, Fuzzy logic, Vector query, and Extended Boolean logic. The aim of information retrieval is to access retrospective information of organization and to share for all users who need information.

2.4. KNOWLEDGE USING Organizations use knowledge for three reasons: 1) Knowledge can be used for determining organization’s work processes and making strategies for sustainable competitive advantage. 2) Knowledge can be used for designing and marketing product. 3) Knowledge plays a critical role of organization’s services quality.

Also, Alavi emphasized that knowledge can be used through three basic mechanisms: Directives that refers to specific set of rules, standards, procedures, and instructions developed through the conversions of the specialist’s tacit knowledge to explicit and integrated knowledge for efficient communication to non-specialist. Organizational routines refer to the development of task performance and coordination patterns, interaction protocols and process specifications that allow individuals to apply and integrate their specialized knowledge without the need to articulate and communicate what they know to others. Self-contained task teams refer to task uncertainty and complexity prevent the specification of directives and organizational routines, teams of individuals with prerequisite knowledge and specialty are formed for problem solving. Like knowledge structuring, knowledge using is also based on information technology. For this reason, if individuals would like to use information effectively, they firstly should be information literacy.

2.5. KNOWLEDGE AUDITING Knowledge auditing means what amount of knowledge can be used in organization’s products, services and processes. This knowledge management life cycle stage refers to the capacity of information processing in organizations. In other words, what amount of information and knowledge are created, shared, stored, and used in organization in a certain time helps us to determine information capacity in organizations.

The knowledge audit provides value when company is doing one or more of the following:
• devising a knowledge-based strategy,

• architecting a knowledge management blueprint or roadmap,

• planning and build a knowledge management system

• planning research and development

• seeking to leverage its ‘people assets’

• facing competition from knowledge intensive competitors that are far ahead on the learning curve

• striving to strengthen its own competitive weakness

• looking for direction for planning a market entry or exist strategy.

Another critical factor for auditing knowledge in organization is measuring intellectual capital, intangibles such as information, knowledge and skills that can be leveraged by an organization to produce an asset of equal or greater importance than land, labor and capital.

CONCLUSION

Information Retrieval is very much helpful in decision making and can be best implement with the models and methods available. Knowledge management application begins with creating knowledge in two forms, tacit and explicit; and goes on knowledge sharing through social and technical communication infrastructure. After sharing, it is needed to structure knowledge at three stages for the retrospective usage: 1-information mapping, 2-information storing and 3-information retrieval. Structured knowledge is ready for using the organization’s products, services and work processes that gains competitive advantage, increases innovative capacity and R&D in organizations. Knowledge auditing is the last stage of knowledge management life cycle in organization as an application. Through this stage, organization can realize the amount of data, information and knowledge by measuring intellectual capital.

REFERENCES

1. AN INTRODUCTION TO DATAMINING
   www.thearling.com/text/dmwhite/dmwhite.htm

2. INFORMATION RETRIEVAL
   www.automationanywhere.com/extract

3. KNOWLEDGE MANAGEMENT
   www.systems-thinking.org/kmgmt/kmgmt.htm

4. KNOWLEDGE MANAGEMENT SOLUTIONS AND DEFINITIONS
   www.cio.com › ... › Knowledge Management (KM) › Tutorial