IJDCST @November Issue- V-1, I-7, SW-93
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Enhanced Multi Attribute Fuzzy Search over XML Data

[1]Ramya Talasila , [2]D.SreeLakshmi ,
[1] Student, PVPSIT, KANURU, VIJAYAWADA, KRISHNA DIST.
[2] Assistant Professor, PVPSIT, KANURU, VIJAYAWADA, KRISHNA DIST.

Abstract: The Efficient state of the art query retrieval systems cannot be implemented on the XML based storage but can implement on the RDBMS databases. Fuzzy type-ahead search is a new information access paradigm for XML based systems though not a new concept for RDBMS based systems. Operations involving the system searching the XML repositories on the fly as the user types in query keywords for producing results. In the XML storage fast querying and result generation is the vital aspect. Prior systems used LCA-based (Lowest Common Ancestors) algorithms for implementing fuzzy type-ahead search and Minimal-Cost Tree based techniques for top-k results over xml data. The minimal-cost tree based approaches are efficient as long as the query keywords are singular or dual utmost. No. of attributes in the keyword for fuzzy query increases Minimal-Cost Tree construction is a computationally expensive process. We propose to use Fuzzy Multiple Attribute Decision Making (FMADM) algorithm involving data conflict resolution based on subjective and objective weighting methods. We intend to support multi attribute based queries over xml data with reduced computations based on the FMADM algorithm. A practical implementation of the proposed system validates our claim.

Keywords: Fuzzy Search, XML, RDBMS, FMADM.

I. INTRODUCTION

The extreme success of web search engines makes keyword search the most popular search model for ordinary users. XML is becoming a standard in data representation, it is desirable to support keyword search in XML database. XML data extraction from multi dimensional. It is a user friendly way to query XML databases since it allows users to pose queries without the knowledge of complex query languages and the database schema. In most systems that incorporate keyword search into relational or XML data, the sole criterion is proximity.

It is argued that in a tree document that the keywords are semantically related if they appeared in a uniquely labeled sub tree of the document. In the work it is improved by introducing an approach that avoids some cases of incorrect results. A result snippet should represent a semantic unit to be self-contained. The fragment between keyword matches in the corresponding XML document as the snippet of this query result that the users will not be able to see that both matches are nested in the tag retailer and thus not able to easily understand that this query result is about an apparel retailer in Texas.

To achieve this in text document search, result snippets often include the document titles. Numerous search tools have been developed to perform keyword searches and locate personal information stored in file systems. Tools usually support some form of ranking for the textual part of the query—similar to what has been done in the Information Retrieval (IR) community—but only consider structure and metadata as filtering conditions. The research community has turned its focus on search over to Personal Information and Data spaces that consist of heterogeneous data collections.

These works focus on IR-style keyword queries and use other system information only to guide the keyword-based search. The contributions of our work include:

a. The problem of generating query result snippets for XML search.
b. Four goals are identified to meet the good query result snippets
c. To address the goals, we identify the significant information in a query result to be selected into the snippet.
d. We can construct a snippet of a given size limit that contains all the significant information, we prove that the decision problem.
e. Generating snippet for XML search has been implemented and tested for its efficiency and effectiveness through experimental studies.

Fuzzy Multi-Attribute Decision Making is a method used to find the optimal alternative from a number of alternatives to certain criteria. It is the core of determining the value of the weights for each attribute. There are three approaches to find the weights of attributes, namely:

- *Approach of subjective*

  The weights are determined based on the subjectivity of decision-makers par.
- *Objective approach*

  The weights are calculated mathematically that ignoring the subjectivity of the decision makers.
- *Approach to the integration between the subjective and objective*

Unified Multi-Dimensional Scoring: We present our unified framework for assigning scores to files based on how closely they match query conditions within different query dimensions. We distinguish three scoring dimensions: content for conditions on the textual content of the files, metadata for conditions on the system information related to the files, and structure for conditions on the directory path to access the file. We use a simplified version of XQuery to express metadata and structure conditions in addition to keyword-based content conditions. Scores across multiple dimensions are unified into a single overall score for ranking of answers. For each query condition, we score files based on the least relaxed form of the condition that each file matches. The scoring along all dimensions is uniformly IDF-based which permits us to meaningfully aggregate multiple single-dimensional scores into a unified multi-dimensional score.

II. RELATED WORK

Inspired by the great success of IR approach on web search, we aim to achieve similar success on XML keyword search, to solve the above three issues without using any schema knowledge. Main challenge we are going to solve is how to extend the keyword search techniques in text databases (IR) to XML databases. The basic data units in text databases searched by users are flat documents. IR systems compute a numeric score for each document and rank the document by this score. In XML databases information is stored in hierarchical tree structures. The statistics is a mathematical science pertaining to the collection, analysis, interpretation or explanation of data. Although keyword search is a subjective problem that different people may have different interpretations on the same keyword query, statistics provides an objective way to distinguish the major search intention(s).

FMADM methods basically involve two phases before to achieve a decision: aggregation and exploitation. Aggregation phase combines the performance ratings for all attributes with respect to each alternative. Exploitation phase ranks the alternatives with respect to the global aggregated performance ratings. The literature contains numerous applications of FMADM to different aspects of selection problems with vague data, propulsion system selection, and advanced manufacturing systems selection.

A multiple attribute decision-making problem considered in this study composes the following elements: Let A= {A1, A2… Am} comprising a finite set of alternatives and moreover let there be a finite set of attributes C= {C1, C2… Ck}, where these attributes are classified as subjective attributes {C1, C2… Cs} and objective attributes {Cs+1, C s+2… Ck}. A decision maker cannot generally specify precise numerical values they can take the form of linguistic variables or fuzzy numbers because

- A decision should be made to experience time pressure and lack of knowledge or data
- Numerous attributes are subjective or intangible owing to being unquantifiable in nature
- Precise quantitative or non-monetary information may not be stated because it is either unavailable or too costly to compute

Fuzzy numbers are very useful in improving information representation and processing in a fuzzy environment. Trapezoidal fuzzy numbers have been used to characterize linguistic labels used in approximate reasoning. Let a fuzzy number A be a special fuzzy subset of a universal set X with membership function, that is a continuous mapping from each element x in X to a real number in the interval [0, 1]. This study presents a fusion method of fuzzy information, which is performed in two phases:

- Making the information uniform
- Computing the collective information

III. EXISTING SYSTEM

Efficient query retrieval systems are for RDBMS systems solely and not for XML based mostly systems. Uses keyword-search system over XML information. A user composes a keyword query, submits it to the system, and retrieves relevant answers. This is often known as try-and-see approach wherever user's limited information regarding the data forces them to be content with limited query results. The try-and-see approach systems don't support users enlarged information domains.

Query results are influenced by minor errors in keywords. Thus an improved system is required that supports users enlarged information domains and additionally robust to minor errors in keywords. Even though this concept is nothing new for RDBMS based systems, this is a new information-access paradigm for XML based systems.

Here, the system searches XML data on the fly as the user types in query keywords. Benefits of the proposed system includes the following

- Auto complete features
- Supports Fuzzy Search over XML Data
- Effective index structures and searching algorithms over XML drives top-k results

Uses the following algorithms and techniques

- LCA-based(Lowest Common Ancestors) or MCT-based(minimum connecting trees) fuzzy type-ahead search algorithms
- Ranking Minimal-Cost Tree based techniques for top-k results

Produces high search efficiency and result quality over XML data storages.

IV. PROPOSED SYSTEM

Prior Systems Use Minimal-Cost Tree based techniques for producing top-k results. Minimal-Cost Tree based approaches are efficient as long as the query keywords are singular or dual utmost. As the number of attributes in the keyword for fuzzy query increases Minimal-Cost Tree construction is a computationally expensive process. So we propose to use Fuzzy Multiple Attribute Decision Making (FMADM) algorithm to support multi attribute based queries at a significantly lesser computations.

FMADM problems are organized as a matrix in which some alternatives evaluate conflicts based on some criteria. For each criterion, we must assign a weight that describes its relativity importance. The best alternative is obtained by the affecting weights vector on decision matrix. There are two categories of weighting methods: subjective and objective methods.

- The subjective methods determine weights solely according to the preference or judgments of decision makers. Then apply some mathematic methods such as the eigenvector method, weighted least square method, and mathematical programming models to calculate overall evaluation of each decision maker.
- The objective methods determine weights by solving mathematical models automatically without any consideration of the decision maker’s preferences.

Based on the Fuzzy Multiple Attribute Decision Making (FMADM) algorithm we intend to support multi attribute based queries over xml data with reduced computations.

V. FUZZY MULTIPLE ATTRIBUTE DECISION MAKING

FMADM methods basically involve two phases before to achieve a decision: aggregation and exploitation. Aggregation phase combines the performance ratings for all attributes with respect to each alternative. Exploitation phase ranks the alternatives with respect to the global aggregated performance ratings. A multiple attribute decision-making problem considered in this study composes the following elements: Let A= {A1, A2… Am} comprising a finite set of alternatives and moreover let there be a finite set of attributes C= {C1, C2… Ck}, where these attributes are classified as subjective attributes {C1, C2… Cs} and objective attributes {Cs+1, C s+2… Ck}.

Let a fuzzy number A be a special fuzzy subset of a universal set X with membership function, that is a continuous mapping from each element x in X to a real number in the interval [0, 1]. Assumes that a trapezoidal fuzzy number is represented by the membership function:

$$\mu_B(x) = \begin{cases} (x - n_1)/(n_2 - n_1), & n_1 \le x \le n_2, \\ 1, & n_2 \le x \le n_3, \\ (x - n_4)/(n_3 - n_4), & n_3 \le x \le n_4, \\ 0, & \text{otherwise.} \end{cases}$$

With n1≤n2≤n3≤n4. The x in interval [n2, n3] yields the maximal grade of μ B (x), which is the most likely value of the evaluation data. The n1 and n4 comprise the lower and upper limits of the available area for the evaluation data, which are used to reflect the fuzziness of the evaluation data.

Making the information uniform:

The expert fuzzy assessments must be converted into a basic linguistic scale. Each assessment value is defined as a fuzzy set on the basic linguistic scale with respect to the assessments made by linguistic labels in the linguistic scale. Let S={s0, s1,…, sT} and V={v0, v1,…, vG} be two linguistic scales. A transformation function θ SV is then defined as

$$\boldsymbol{\theta_{SV} : S \rightarrow F(V),}$$

$$\boldsymbol{\theta_{SV}(si)=\{(u_{ij}, v_j)/j\in\{0, 1,\ldots, G\}\} \text{ for } si\in S,}$$

$$\boldsymbol{uij = {}_{x}^{max}\min\{\mu_{si}(x),\mu_{vj}(x)\},}$$

where F(V) denotes the set of fuzzy sets defined in V. Regarding the fuzzy assessments assessed by approximate numerical values the transformation function also appropriately implemented to converting the standardized fuzzy assessments. The converted information provided by experts for each alternative with respect to attribute is defined as the fuzzy set on the basic linguistic scale V. The collective performance rating of an alternative is then obtained by aggregating these fuzzy sets.

The OWA operators can be provided for aggregating the attributes associated with some linguistic fuzzy quantifiers such as 'as many as possible', 'average', 'most', almost all' 'at least half'. In decision-making fuzzy quantifiers are used to indicate a fusion strategy for guiding the process of aggregating expert opinions. This method enables the experts' fuzzy assessments with the linguistic and numerical scales can be considered in the aggregation process.

ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Fig 2: Algorithm for FMADM.

Let S be an appropriate linguistic scale chosen by the committee be used for the qualitative assessment versus subjective attributes. For handling the fuzzy information, all the linguistic weighted ratings are transformed into their corresponding fuzzy numbers with linguistic scale S. Each weighted rating can be defined as a fuzzy set on the basic linguistic scale. To ensure compatibility among the various numerical scales, all the estimated values must be converted into a comparable scale. Let a trapezoidal fuzzy number Wt=(at, bt, ct, dt) be used to denote the importance grade for objective attribute Ct assigned by experts. let H_{it} =(e_{it}, g_{it}, h_{it}, l_{it}) represent a positive trapezoidal fuzzy number representing the estimated performance rating for an alternative Ai with respect to an objective attribute Ct.

VI. EXPERIMENTAL RESULTS

The application of the proposed selection method to the name of the firm is not revealed for reasons of confidentiality. A feasibility evaluation of Optional Operation System for the bicycle industry, and the data was taken from a study entitled 'Fuzzy multi-attribute decision-making'. A few years ago the case firm decided that it needed a flexible manufacturing system (FMS) that would allow a customized bike. Each bike consists of 11 subsystems including a frame, derailleur shifters, suspension fork, brokers, hubs and rims, pedals, tires, stem, handle bar, saddle, and seat post. Each subsystem includes several models among which customers can select. The effects of adopting a FMS include both quantitative

Input: Set of data

Output: Search results

Step 1: Enter into hierarchy model in the form of data sets

A0, A1, A2,………………………….An.

Step 2: Verify each attribute Domain results

Step 3: Construct Matrix Representation.

Ai(i=1,2,3,….),h_{ij} i=1,2,3,…. J = 1,2,3,……

Step 4: Construct m-dimensional fuzzy attribute evolution process for each attribute.

Step 5: Compressive weights of all alternative attributes

Step 6: Comprehensive ranking of all attributes.

Step 5: Search results.

and qualitative effects. While objective attributes are used to evaluate the quantitative effects of FMS. Table 1 lists the properties of these attributes, including attribute type and assessment type

Subjective attribute	**Objective attribute**
C1：Process flexibility	C5：Required floor space (ft2×1,000)
C2：Product quality	C6：Capacity (unit×1,000)
C3：Learning	C7：Lead time (hr)
C4：Exposure to labor unrest	C8：Purchase cost ($×10,000)

Table 1: The FMS selection attributes

With respect to each of the objective attributes (C5, C6, C7 and C8) based on the evaluation information provided by management. The performance rating and important grade for each FMS are identified by a group of experts. The standardized performance ratings for each alternative versus the objective attributes can generate. For alternative A1 with respect to C5, the performance rating of (5, 6, 6.5, 7) can be transformed into the standardized performance rating. With respect to the detailed analysis of evaluation results such as competing FMS alternatives effects of FMS. The decision-making process will be completed if experts accept the evaluation results. Experts can modify their opinions step by step through the collection of additional information or modify the linguistic fuzzy quantifier until a consistent decision is obtained.

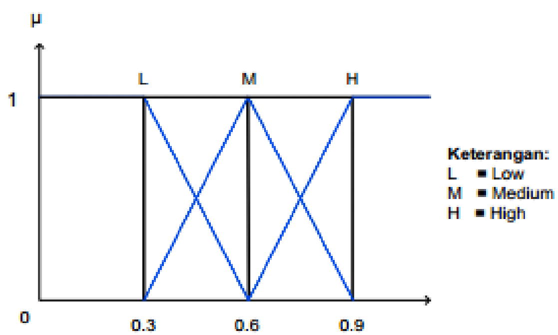

Fig. 1.Overview of fuzzy linguistic variables

VII. CONCLUSION

Operations involving the system searching the XML repositories on the fly as the user types in query keywords for producing results. Prior systems used LCA-based algorithms for implementing fuzzy type-ahead search and Minimal-Cost Tree based techniques for top-k results over xml data. As the number of attributes in the keyword for fuzzy query increases Minimal-Cost Tree construction is a computationally expensive process. We propose to use Fuzzy Multiple Attribute Decision Making (FMADM) algorithm involving data conflict resolution based on subjective and objective weighting methods. Our experimental result shows efficient data retrieval techniques on data efficiency. A fuzzy multiple attributes decision-making scenario was modeled to solve the AMT evaluation problem. We also present a new fusion approach of fuzzy information. According to decision-makers' attitude a linguistic fuzzy quantifier chosen by the manager of the decision problem. The proposed method enables the decision-makers to incorporate and aggregate fuzzy information provided for multiple attributes. As further improvement our proposed work it will be provide more accessing device specification through data aggregations present in data extraction.

VIII. REFERENCE

[1] S Jianhua Feng, Guoliang Li, "Efficient Fuzzy Type-Ahead Search in XML Data," Proc. IEEE transactions on Knowledge and Data Engineering, VOL. 24, NO. 5, MAY 2012.

[2] Shian-Jong Chuu., 2001, "A Fuzzy Multiple attributes Decision-Making for the Evaluation of Advanced Manufacturing Technology", pp.217-242.

[3] Widayanti-Deni, Oka-Sudana and Arya-Sasmita , "Analysis and Implementation Fuzzy Multi-Attribute Decision Making SAW Method for Selection of High Achieving Students in Faculty Level," Proc. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013.

[4] P.Santhi, K.Kiran Kumar, "Data Extraction Based on Multi Dimensional Data Attributes Using Top-K Results," Proc. NTERNATIONAL JOURNAL FOR

DEVELOPMENT OF COMPUTER SCIENCE & TECHNOLOGY , October Issue- V-1, I-6, 2013.

[5] Buckley J.J. (1984) Fuzzy Sets and Systems, 13 (1), 25-37.

[6] Feng C.-M. and Wang R.T. (2001) Transport Reviews, 21 (4), 449-467.

[7] D. Zhang, Y.M. Chee, A. Mondal, A.K.H. Tung, and M. Kitsuregawa, “Keyword Search in Spatial Databases: Towards Searching by Document,” Proc. Int’l Conf. Data Eng. (ICDE), pp. 688-699, 2009.

[8] G. Koutrika, Z.M. Zadeh, and H. Garcia-Molina, “Data Clouds: Summarizing Keyword Search Results over Structured Data,” Proc. Int’l Conf. Extending Database Technology: Advances in Database Technology (EDBT), pp. 391-402, 2009.

[9] L. Qin, J.X. Yu, and L. Chang, “Keyword Search in Databases: The Power of Rdbms,” Proc. ACM SIGMOD Int’l Conf. Management of Data, pp. 681-694, 2009.

[10] T. Tran, H. Wang, S. Rudolph, and P. Cimiano, “Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data,” Proc. Int’l Conf. Data Eng. (ICDE), pp. 405-416, 2009.