



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

Boolean Retrieval for Medical datas

ABSTRACT:

Extended Boolean retrieval (EBR) models were proposed nearly three decades ago, but have had little practical impact, despite their significant advantages compared to either ranked keyword or pure Boolean retrieval. In particular, EBR models produce meaningful rankings; their query model allows the representation of complex concepts in an and-or format; and they are scrutable, in that the score assigned to a document depends solely on the content of that document, unaffected by any collection statistics or other external factors. These characteristics make EBR models attractive in domains typified by medical and legal searching, where the emphasis is on iterative development of reproducible complex queries of dozens or even hundreds of terms. However, EBR is much more computationally expensive than the alternatives. We consider the implementation of the p-norm approach to EBR, and demonstrate that ideas used in the max-score and wand exact optimization techniques for ranked keyword retrieval can be adapted to allow selective bypass of documents via a low-cost screening process for this and similar retrieval models. We also propose term independent bounds that are able to further reduce the number of score calculations for short, simple queries under the extended Boolean retrieval model. Together, these methods yield an overall saving



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

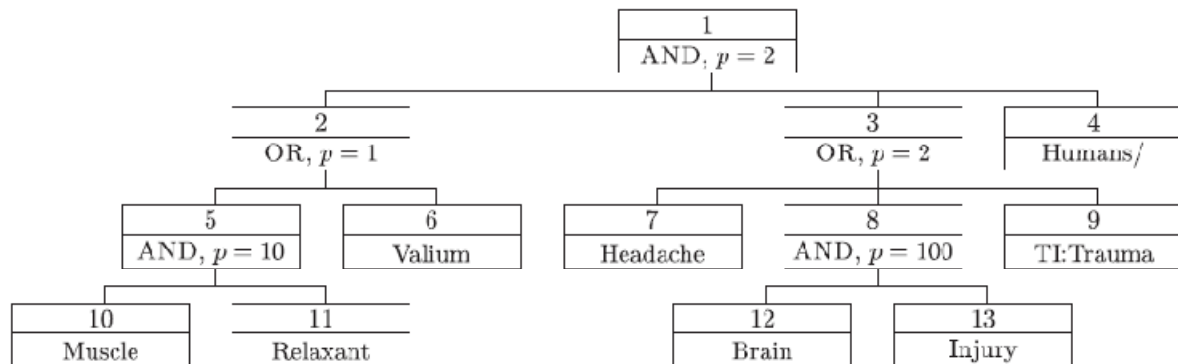
Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

from 50 to 80 percent of the evaluation cost on test queries drawn from biomedical search.

SYSTEM ARCHITECTURE:



EXISTING SYSTEM:

In our Existing System, A significant amount of work has been devoted to the evaluation of top-k queries in databases. Provide a survey of the research on top-k queries on relational databases. This line of work typically handles the aggregation of attribute values of objects in the case where the attribute values lie in different sources or in a single source. For example, Bruno etc. Consider the problem of ordering a set of restaurants by distance and price. They present an optimal



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

sequence of random or sequential accesses on the sources (e.g., Zagat for price and Mapquest for distance) in order to compute the top- k restaurants. Since the concept of top-k is typically a heuristic itself for locating the most interesting items in the database, Theobald et al. Describe a framework for generating an approximate top-k answer, with some probabilistic guarantees. In our work, we use the same idea; the main and crucial difference is that we only have “random access” to the underlying database (i.e., through querying), and no “sorted access.” Theobald et al. assumed that at least one source provides “sorted access” to the underlying content.

DISADVANTAGES OF EXISTING SYSTEM:

The queries are often much more complex, involving dozens or even hundreds of terms; there is a great deal of reformulation and reevaluation and the user evaluation process typically involves hundreds or thousands of answer documents rather than a mere handful.

PROPOSED SYSTEM:

We present a scoring method for EBR models that decouples document scoring from the inverted list evaluation strategy, allowing free optimization of the latter.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

The method incurs partial sorting overhead, but, at the same time, reduces the number of query nodes that have to be considered in order to score a document. We show experimentally that overall the gains are greater than the costs. We adopt ideas from the max-score and wand algorithms and generalize them to be applicable in the context of models with hierarchical query specifications and monotonic score aggregation functions. Further, we show that the p-norm EBR model is an instance of such models and that performance gains can be attained that are similar to the ones available when evaluating ranked queries. Term-independent bounds are proposed, which complement the bounds obtained from max-score. Taken alone, term-independent bounds can be employed in the wand algorithm, also reducing the number of score evaluations. Further, in conjunction with the adaption of max-score, this novel heuristic is able to short-circuit the scoring of documents.

ADVANTAGES OF PROPOSED SYSTEM:

Complex information need descriptions: Boolean queries can be used to express complex concepts.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

Composability & Re-use: Boolean filters and concepts can be recombined into larger query tree structures.

Reproducibility: Scoring of a document only depends on the document itself, not statistics of the whole collection, and can be reproduced with knowledge of the query.

Scrutability: Properties of retrieved documents can be understood simply by inspection of the query.

Strictness: Strict inclusion and exclusion criteria are inherently supported, for instance, based on metadata.

MODULES:

1. Using Boolean Condition (AND)
2. Using Boolean Condition (OR)
3. Using Boolean Condition (NOT)
4. Top k-Query Search



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

MODULES DESCRIPTION:

Using AND Condition:

We define the novel problem of applying ranking on top of sources with no ranking capabilities by exploiting their query interface.

For instance, if the user query is $Q = [\text{anemia}, \text{diabetes}, \text{sclerosis}]$, then we can submit to the data source queries $q_1 = [\text{anemia AND diabetes AND sclerosis}]$, $q_2 = [\text{anemia AND diabetes AND NOT sclerosis}]$, $q_3 = [\text{diabetes OR sclerosis}]$, and so on. The returned results are guaranteed to match the Boolean conditions but the documents are not expected to be ranked in any useful manner.

Using OR Condition:

We describe sampling strategies for estimating the relevance of the documents retrieved by different keyword queries. We present a static sampling approach and a dynamic sampling approach that simultaneously executes the query, estimates the parameters required for efficient query execution, and compensates for the biases in the sampling process. For instance, if the user query is $Q = [\text{anemia}, \text{diabetes}, \text{sclerosis}]$, then we can submit to the data source queries $q_1 = [\text{anemia AND diabetes AND sclerosis}]$, $q_2 = [\text{anemia AND diabetes AND NOT sclerosis}]$, $q_3 = [\text{diabetes OR sclerosis}]$, and so on. The returned results are guaranteed to match the



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

Boolean conditions but the documents are not expected to be ranked in any useful manner.

Using NOT Condition:

We present algorithms that, given a user confidence input, retrieve a minimal number of results from the source through submitting high selectivity (conjunctive) queries, so that the user's confidence requirement is satisfied. For instance, if the user query is $Q = [\text{anemia}, \text{diabetes}, \text{sclerosis}]$, then we can submit to the data source queries $q_1 = [\text{anemia AND diabetes AND sclerosis}]$, $q_2 = [\text{anemia AND diabetes AND NOT sclerosis}]$, $q_3 = [\text{diabetes OR sclerosis}]$, and so on.

The returned results are guaranteed to match the Boolean conditions but the documents are not expected to be ranked in any useful manner.

Top K-Query Search:

Our overall goal is to figure out during our querying process, how many of the top-k relevant documents we have retrieved and how many are still unretrieved in the database. Unfortunately, we cannot be absolutely certain about these numbers unless we retrieve and score all documents: an expensive operation. Alternatively, we can build a probabilistic model of score distributions and examine,



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

probabilistically, how many good documents are still not retrieved. We describe our approach here.

SYSTEM CONFIGURATION:-

HARDWARE REQUIREMENTS:-

- ✓ Processor - Pentium –III
- ✓ Speed - 1.1 Ghz
- ✓ RAM - 256 MB(min)
- ✓ Hard Disk - 20 GB
- ✓ Floppy Drive - 1.44 MB
- ✓ Key Board - Standard Windows Keyboard
- ✓ Mouse - Two or Three Button Mouse
- ✓ Monitor - SVGA



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

SOFTWARE REQUIREMENTS:-

- ❖ Operating System : Windows95/98/2000/XP
- ❖ Application Server : Tomcat5.0/6.X
- ❖ Front End : HTML, Java, JSP
- ❖ Script : JavaScript.
- ❖ Server side Script : Java Server Pages.
- ❖ Database : MYSQL