Answer yes/no queries in search engines

Chunbin Lin
University of California, San Diego
chunbinlin@cs.ucsd.edu

ABSTRACT

Keyword search is a proven and widely popular mechanism for query processing in search engines that alleviates users from understanding data structures and learning query languages. To support efficient keyword search, inverted indices are widely employed. An inverted index is a collection of inverted lists. Each keyword is associated with an inverted list, which contains all the document ids including the keyword. To save space, inverted list compression schemes are employed such as VB, PforDelta, Simple16, GroupVB, Simple8b, SIMDPforDelta and SIMDBP128. The returned answers for keyword queries are a list of ranked documents containing all the query keywords.

By analyzing the query logs in search engines, we observe that most keyword queries are *yes/no queries*, which expected to get a decisional answer (yes or no). For example, consider query "is Obama the 42nd president of the United States", the expected answer is just a simple "No" (may be also some further evidences). However, existing search engines only return a list of ranked documents containing the keywords within the query. Users have to go through those returned documents in order to figure out one simple answer "No". We argue that handling yes/no queries in search engines will enhance the power of current search engines and will significantly improve the users' search experiences.

Yes/no queries Yes/no queries can be divided into two categories *objective yes/no queries (OYN)* and *subjective yes/no queries (SYN)*. The answers of OYN queries are simple "Yes" or "No" with evidences, which can be obtained based on existing facts. For example the answer to the OYN question "is universal studios hollywood open now?" is "Yes" if the following two facts hold. (1) The open time of Universal Studios Hollywood is 9AMÍC10PM from Monday to Friday, and (2) it is 9:30AM Monday now. Otherwise, the answer becomes "No". The answers of SYN queries are $(Yes, \mathcal{P}_y)$ and $(No, \mathcal{P}_n)$ pairs with evidences, where $\mathcal{P}_y$ and $\mathcal{P}_n$ are the probability of the answers to be "yes" and "no" respectively. No facts can be directly utilized to answer such queries, instead users' subjective opinions can be used. For example the answer to the SYN question "is ios better than android" is "(Yes, 55%), (No, 45%)" based on analyzing users' comments, tweets and blogs.

Architecture of processing yes/no queries. Figure 1 shows the architecture of processing yes/no queries. The *Query Analyzer* analyzes the query to identify the type. The *Context Manager* collects users' time-spatial information and also relevant documents from search engines via keyword search. Time-spatial information is served as filter condition, while keyword search returns only relevant context of queries. For example, in order to answer the query "is it rainy today", the following three information should be obtained: (1) today's date, e.g., 8/1/2016, which is the time information, (2) user's location, e.g., New York, which is the spatial information, and (3) the weather of New York on 8/1/2016. The *OYN/SYN Processor* processes the OYN and SYN queries based on the corresponding context and query types, and returns answers in corresponding format.

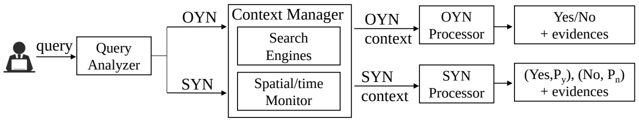
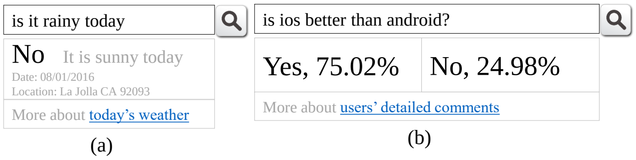


Figure 1: Architecture of processing yes/no queries.

Results visualization Figure 2(a) shows an example visualization for the OYN query "is it rainy today", while Figure 2(b) visualizes an example result of the SYN query "is ios better than android". The outputs give the decisional answers. Users can also see more details to see the evidences.


is it rainy today

No It is sunny today
Date: 08/01/2016
Location: La Jolla CA 92093

More about today's weather

(a)

is ios better than android?

Yes, 75.02% No, 24.98%

More about users' detailed comments

(b)

Figure 2: Example answers of a yes/no query.

Remark Answering yes/no queries in search engines is different from the existing question/answer systems, which process natural language questions by using existing evidences. Nevertheless, we propose to answer yes/no queries based on the keyword search results plus time-spatial information. This paper paves a path for the future direction of the next generation search engines, i.e., answer yes/no queries in search engines.