

Title	電子化辞書を利用した、概念に基づくクエリーの拡張に関する研究
Author(s)	太田, 千晶
Citation	
Issue Date	1998-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1113
Rights	
Description	Supervisor:奥村 学, 情報科学研究科, 修士

Concept Based Query Expansion using Machine-Readable Dictionaries

Chiaki Ohta

School of Information Science,
Japan Advanced Institute of Science and Technology

February 14, 1998

Keywords: Query Expansion, Concept, Information Retrieval(IR),
Machine-Readable Dictionaries.

In this paper, we propose an automatic query expansion method using “concept” described in large lexical database.

Because most of information around us are made into the electronic data and the computer have come into wide use recently, many people would expect not only the high skill of information retrieval and also the complete support system for the end users on IR system. Specially the end users had many chances to access these large electronic database(e.x. World Wide Web), and then the difficulties about IR system had disclosure by reason of the users themselves.

In most of information retrieval systems, relevant documents are usually retrieved by simply matching words in an initial query and documents. So in most cases, it is impossible to retrieve all relevant documents perfectly by the user's initial query.

The query expansion can solve this problems. The query expansion system can expands a user's query into some useful query terms for IR, and make some improvements of the precision consequently.

We'd built a query expansion system using concepts and various concept relations described in EDR, a large scale electronic lexical knowledge database.

We are sure that we can expand the user's query with reliable sense relevance by using “concept” in EDR, and get some reliable relevance terms after the expansion.

we first propose two query expansion methods below.

(method 1) a method using relations between noun and verb

By using the records in CPT.DIC of EDR, we expand the nominal(verbal) concept of an initial query into the another verbal(nominal) concept that co-occur the concept of an initial query plausibly. Then we create an effective query by combining these expanded words and the synonyms of an initial query. For example, we make queries of this type “[*theinitialqueryterm*] AND [*expandedqueryterm*]” using co-occur concepts. We expect that these query terms play a important role that they obtain only the right document from the all documents included initial query.

(method 2) a method using “Concept Explanation”

At first, we search “Concept Explanation” that include 2 initial query terms. Second we obtain the concept of the last “Concept Explanation” as relevant concept.

We can obtain some relevant terms from this relevant concept, and these terms are different from another relevant terms from the usual expansion method(with same, high, or low concept).

We also consider that word sense disambiguation of an initial query may bring about the positive influence to improve precision in document retrieval. Therefore we propose three automatic word sense disambiguation methods.

Experiments

Experiments are carried out using “Bench Mark for Japanese IR Systems Ver.1.0(BMIR-J1)” : including 600 documents, 40 queries and list of their relevant document number.

The effectiveness of document retrieval is evaluated by recall and precision.

(method 1)

We could only a slight improvement on precision by this (method 1). However our analysis showed that 9 queries’s case between all 40 queries could obtain much improvement compared with unexpanded query(= Same sense words only) using this method.

And we pay attention to the co-occurrence of noun and verb, then we try another experiments : a document is decided that is right document for the query only if the distance noun and verb is 2 sentences or under. Then we could much improvement of precision at almost recall points.

(method 2)

We could obtain some relevant concept of 20 queries between all 40 queries and the number of obtained concept is 10, and number of relevant terms is 12.

Then using our (method 2), we could some queries which precision improved more than unexpected query.

(disambiguation)

At first we experimented with the queries that all ambiguous terms are disambiguated by hand. In this result, the query with disambiguation by hand marked the highest precision, and query with disambiguation by our method (this method are combined all our three methods) are 1% - 3% than the query without disambiguation.

Our combined disambiguation method could really disambiguate 50% of all ambiguous terms.

Our future works are described below.

- We need to know the sentence structure of query and Concept Explanation and retrieval documents if we want to do exact term matching or other element on expansion process.
- We have to experiment again with a more large scale Bench Mark.