

Title	XML検索におけるデバッグ支援システム
Author(s)	内田, 隆彦
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1989
Rights	
Description	Supervisor: 田島 敬史, 情報科学研究科, 修士

A XML Query Debugging System

Takahiko Uchida (410017)

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

February 9, 2006

Keywords: XPath,debugging,query.

1 Abstract

When users issue XPath queries against XML data, it often happens that they cannot get what they wanted because of some errors in the queries. Such errors include typos and misunderstanding of the structure of the target XML data. In this paper, we propose a system to help users debugging their XPath queries. In our system, when a user gets a query answer that is not what they wanted, the user tells the system how that answer disagrees with the users intention. Then, the system finds queries that are similar to the original queries and that solves the disagreement with the users intention, and the system shows those queries in the order of the probability of being the query the user was intending, which is computed based on the statistics information.

2 Outline of system

When the system that describes in the preceding chapter is developed, it is necessary to consider the following three points.

1. Is it made the specification of the user what respect of the inquiry result contradicts the intention very?

2. How is the inquiry type group that cancels the problem in the inquiry result that the neighborhood to a first inquiry type, and the user specified found?
3. How is "Possibility of each inquiry type is an inquiry type that the user wanted to write originally" necessary to do those inquiry types in the ranking requested?

Hereafter, it explains these three points in detail to order.

2.1 Method of specifying point in contradiction to intention

Here, where of the inquiry result contradicts the intention will be specified in the form of either of the user the following.

- Data that this data is not included though should be included in the inquiry result is specified.
- Data that this data is included though should not be included in the inquiry result is specified.
- It is specified that it is not sure to become an empty set though the inquiry result was an empty set.

2.2 Discovery of inquiry type that becomes candidate

Here, it thinks about the following things as a mistake with the possibility that the user violates it, it thinks about the thing that can become an inquiry type that the user wrote first due to the following mistakes (Or, those two or more combinations), and it is assumed to be a candidate solution to meet the requirement that the user specified in them.

2.3 Method of ranking of inquiry type that becomes candidate

It is preferable to do the ranking from the one with high possibility that is the inquiry that the user wanted to write sequentially and to display these candidates when two or more inquiry types become candidates since the inquiry type that becomes a candidate by the above-mentioned method is

requested. In this research, the technique for doing the candidate solution in the ranking based on it is developed for such a probability.

When the inquiry type that the user wrote was Q_0 , the inquiry type that wanted to write really is requested by using the theorem of Bayes from the probability that mistakes the probability that is Q_1 to write probability and inquiry type Q_1 that the user wants to execute inquiry Q_1 and writes Q_0 in the technique of this research. Moreover, the point that it is an inquiry type with the feature that seems to be low the possibility near Q_1 and Q_0 but to write such an inquiry type well simply by an inquiry type that the probability that Q_0 writes due to a mere mistake seems to be high it or a mere mistake when the probability written Q_0 by mistake to write inquiry type Q_1 is requested it is considered.

3 Problem in summary and the future

In this thesis, when it proposed the outline of the system that supported the debugging of the inquiry type when the user inquired XML by XPath, and an inquiry type wrong of the user was written especially, the candidate who seemed it even though the inquiry type that had wanted to write originally was shown about a basic idea of the technique for doing the ranking in order in which the probability seemed to be high.

However, it will be necessary to examine the following problems the stage it where a basic idea was shown, and to achieve an actual system at the present stage in the future.

- Detailed, more formal definition of ranking calculation.
- How is probability $i(Q)$ to want to execute the inquiry with the user given though it described in the text?
- How is the probability of a variety of kinds of mistakes that the user violates those mistakes given?
- The one other than enumerated mistake by three kinds of and enhancing of mistake, "and", and "or" of "/" and "//" this time to mistake etc. for instance .

- Implementation
- Examination how effectiveness of proposal system should be evaluated .

Especially, reactive time until the ranking can be presented to the user becomes very long when the candidate inquiry type increases, and it seems that convenience becomes a bad system about the mounting technique in requesting all the inquiry types that agree with the intention that getatable according to the kind of the mistake given from the inquiry type that the user wrote, and user specified as a very simple mounting method, figuring out the odds about them, and doing the ranking. Therefore, it seems that the technology of the top-k inquiry that is narrowed only to the solution candidate that some techniques are used, and the possibility ranking high by the first ranking is high and calculated is needed. Therefore, the mounting technique is scheduled to be examined especially in the future.