B. Chang et al. (Eds.) (2010). Joint Proceedings of the Work-in-Progress Poster and Invited Young Researcher Symposium
at the 18th International Conference on Computers in Education. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.

Development of IR Tool for Tree-Structured MathML-based Mathematical Descriptions

Tetsuya WATANABE of Author A[a*] & Yoshinori MIYAZAKI of Author B[b]
[a]*Graduate School of Informatics, Shizuoka University, Japan*
[b]*Faculty of Informatics, Shizuoka University, Japan*
*gs09061@s.inf.shizuoka.ac.jp

Abstract: The quantity of Web contents including math has been skyrocketing in recent years, such as Wikipedia articles and BBS focusing on math. Some pieces of previous research have dealt with the development of IR systems targeting MathML-based math expressions. They are, however, still developing in terms of lack of fuzzy search functions or low hit rates. One of the authors in ICCE2008 proposed the IR tool enjoying a fuzzy search function, by adopting regular expressions used in MySQL. In this study, it is our objective to additionally propose a "tree structure" algorithm for the fuzzy search function with better precisions.

Keywords: IR system for math, MathML, fuzzy search, tree structure

Introduction

In this study, it is our objective to propose development of IR systems targeting MathML-based[1] math expressions.[2] is advocated the IR systems of the algorithm that used the regular expression in ICCE2008. Since MathML is an XML-based markup language, it has tree structures by nature.
It is our objective to propose a "tree structure" algorithm for the fuzzy search function with better precisions than [3], and upgrade the interface for a fast input of math expressions being developed by another group in the lab.

1. Previous Research

[4], [5], and [6] are similar to our study although the developed systems are not found completely satisfactory in either that:

- math structure is not fully considered,
- indexing tag information is too rigid to realize fuzzy searching, or
- only partial implementation has been made.

[7] outlines sophisticated IR system for MathML-based math expressions. It further proposes the interface for creating queries combining text- and math-expression editors for fuzzy search using wildcards while there is no description regarding its implementation or results of experiments.
[8] is unique in proposing an IR system by incorporating math expressions in "extended" MathML-formats for better grasping their (mathematical) meanings. Likewise, no results of implementation or experiments are shown.

2. Implementation

2.1 Outline of retrieval with Tree Structure

Since MathML is an XML-based markup language, it has tree structures by nature. Then, implementing a retrieving algorithm with tree structure could be taken for granted. DOM(Document Object Model) format is used for tree-structuring[1].

2.2 Modification of method

In [3], the wildcard "*" use to mean "arbitrary (single) letter or variable(TypeA)". But in this study, use to mean "a string with arbitrary length(TypeB)"(Fig.1).

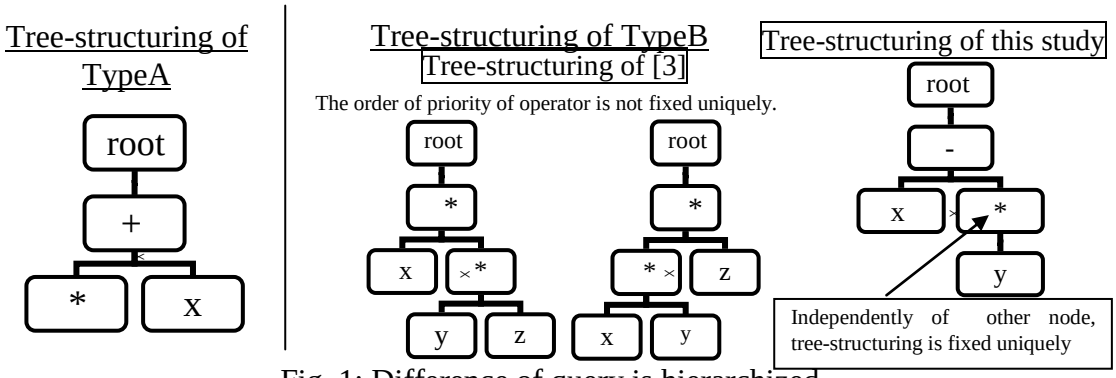

Fig. 1: Difference of query is hierarchized

2.3 Retrieving Algorithm

The wildcard use to part punctuation, and do AND Retrieving and check the hierarchy (Fig.2).

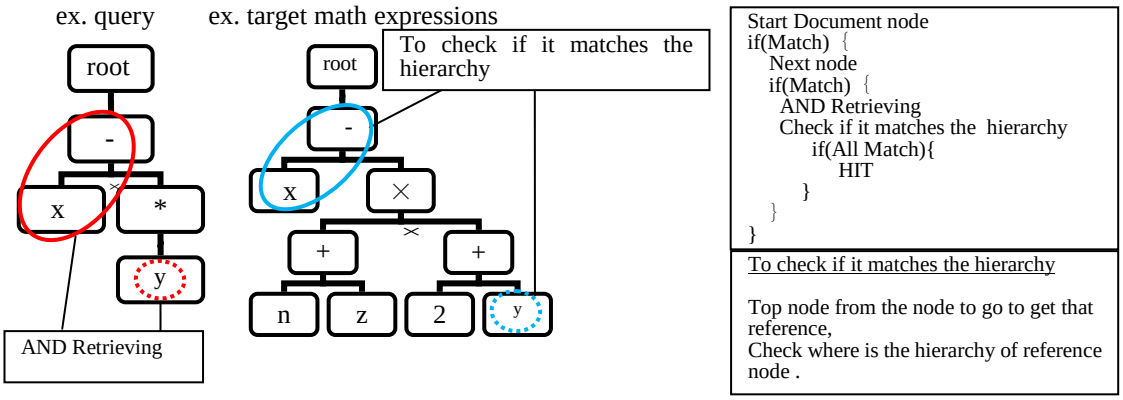

Fig.2:Retrieving Algorithm

3. Language, Style and Content

3.1 Outline of Experiments

[1] DOM enables to treat data described by XML as having tree structures.

IRs were attempted for 1,000 MathML data sampled from [9], with regular expression algorithm and tree structure algorithm, separately. The results of the experiments are shown in Table 1. IR1 to IR3 in the table are the examples of information retrieval, or

IR1: $\cos^2(z)$, IR2: $\cos(z)+*\cos(z)$,, IR3: $\frac{*}{*+*}$

respectively. The numbers of math expressions hit by the selected algorithm are given in the table.

Table 1: Results of Experiments

Retrieval / Algorithm	IR1	IR2	IR3
Tree Structure(TypeA)	**9**	**0**	**67**
Tree Structure(TypeB)	**9**	**1**	**67**
Correct numbers	**9**	**1**	**67**

3.2 Discussions

This time, the specification of structured query [3] to change. We can hit the math expressions that in [3] can hit . And, in this study can hit math expressions like a IR2.

4. Upgrade the interface for a fast input of math expressions

We upgrade the interface for a fast input of math expressions being developed by another group in the lab[10]. Details will be announced on the day.

5. Summary and Future Plan

In this study, it is our objective to propose a “tree structure” algorithm for the fuzzy search function with better precisions than [3], and upgrade the interface. Our future plan is increase in data set. And adding Highlighting.

References

[1] MathML: Mathematical Markup Language Version 2.0 (Second Edition) (2003), http://www.w3.org/TR/MathML2/.
[2] Y. Miyazaki, Y. Iguchi, “Development of Information Retrieval Tool for MathML-based Math Expressions”, Proceedings of the 16th International Conference on Computers in Education (ICCE2008), pp. 419-426 (2008).
[3] T.Watanabe, Y.Miyazaki, Development of Search Engine for MathML-based Mathematical Descriptions, 34th Japanese Society for Information and Systems in Education , pp. 452-453 (2009)
[4] T. Nakanishi, S. Kishimoto, M. Murakata, T. Otsuka, T. Sakurai, and T. Kitagawa, “An Implementation Method of Composite Association Retrieval System for Data of Mathematical Formulas”, DBSJ Letters, Vol. 4, No. 1, pp. 29-32 (2005).
[5] H. Hashimoto, Y. Hijikata, and S. Nishida, “A Survey of index formats for the search of MathML objects”, IPSJ SIG Notes 2007-DBS-142-(10), Vol.2007, No.54(20070531), pp. 55-59 (2007).
[6] K. Otagiri, T. Murata, “Search of Mathematical Formulas Using MathML”, 22nd Annual Conference of the Japanese Society for Artificial Intelligence, (2008).
[7] R. Munavalli, R. Miner, “MathFinder: A Math-Aware Search Engine”, ACM, p 735, ISBN: 1-59593-369-7, (2006).
[8] B. R. Miller, A. Youssef, “Augmenting Presentation MathML for Search”, Intelligent Computer Mathematics, 9th International Conference (AISC 2008), pp. 536-542 (2008).
[9] The Wolfram Functions Site: http://functions.wolfram.com/.
[10] M. Ogasawara, Y. Miyazaki, “Development of Supporting Tool for Making Mathematical Expressions on Blog System That Uses Java Applet”, 34th Annual Conference of IPSJ, pp.454-455 (2009).