SIMILARITY SEARCH FOR MATHEMATICS

MICHÁL RŮŽIČKA, PETR SOJK AND MARTIN LÍŠKA

INTRODUCTION

Masaryk University (MU) has entered the area of MIR during the development of the Czech Digital Mathematics Libraries (CDML) project. CDML digital mathematical libraries (DMLLs) are specific especially in handling of formula.

MU has partnered in the development of the European Digital Mathematics Library (EDML), and indexed math formula as one of the math specific features. We have also paid attention to the user interface aspect of the EdML project in the query are rendered at the same time as the query processing.

EDML with Math Indexing and Searcher (MIS) is the first digital library collecting humangenerated POSEs, which supports math formulas in a full way.

Our MMU team has been participating in MIR, math information retrieval, since their introduction at ICMIR-10. This year we have tried three new approaches: structural unification, new querying strategies and new canonicalization procedures.

In our research we found that the similarity increase is still very low and has negative impact on precision. MIR's A1 truth allowed us to compare effects of canonicalization and different querying strategies.

NEW STRUCTURAL UNIFICATION

An important feature that we reused in NTIC-11 was the ability to normalize (deconstruct) math formulas. We integrated with our WebMath system and structural unification is done during indexing of the document. To our best knowledge, the process has not been used in a full scale application. The unification performed, according to MathML, tree layers.

\[
\frac{a^2 + b^2}{c^2} = \frac{a + \sqrt{c^2 - b^2}}{c}
\]

The MathML Unificator tool is now integrated with our WebMath system and structural unification is done during indexing of the document. As a query processing, the unification is also performed on the queries.

NEW FEATURES

We have built several dozen indices for the main and Web Math with different features and configurations enabled. Every raw in the table corresponds to the particular combination of features. We queried each index with full 3D topics from NTIC-11 with 11 different querying strategies (columns in the table). This gave us 360 results describing the performance of each particular combination. We used MAP and TREC-style metrics for evaluation against ground truth.

The results show that a common structure can be recognized in the search results from the original query.

NEW FUTURE EXPANSION

These modifications are only applicable on multi-word keywords of the original query.

Original query

Formulas: \( a, b, c \)

Keywords: \( a \)

Example of expansion for multi-word keywords individual words are used instead of the original multi-word keyword.

Example: \( a \)

Keywords: \( a \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)

Example of the use of the whole phrase and keywords.

Example: \( a \) and \( b \)

Keywords: \( a \) and \( b \)