

修士論文の和文要旨

大学院 電気通信学 研究科 博士前期課程		情報工学	専攻
氏名	椎名 正樹	学籍番号 0531011	
論文題目	数式を含む文書の検索方式		
<p>要 旨</p> <p>数式データを記述する標準的フォーマットである OpenMath や MathML の登場により、数学的なデータをより豊富かつ再利用可能な形で含む文書の作成が可能になってきた。これらフォーマットによるデータに対する検索が実現すれば、詳細な数学的内容による検索が可能になると考えられる。これが実現すると、数式しか手がかりがない場合でも、それをキーとして検索することで、探している話題を含む文書を大量の文書集合の中から効率よく見つけ出すことができ、便利である。</p> <p>本研究ではこれを受けて、OpenMath や MathML による数式と、本文の両方に対する条件から文書を検索できる文書検索方式を考案、実験的に実装した。これは文書の本文は文字列、数式は S 式と呼ばれるデータに変換してデータベースに登録することで検索するので、「S 式検索システム」と呼んでいる。</p> <p>本方式では文章に対する単語による検索と、数式に関する条件による検索を併用することによって、それぞれの単体による検索より精度良く検索できる。加えて、数式に関してはその木構造も条件に指定することができるので、単なる単語による検索よりもさらに精度を向上できる。更に、構造に関する条件を満たすものだけを検索結果として提示するのではなく、それを上位に提示し、より弱い条件を満たすものを捨てるのではなく、下位に提示することで、構造に対する要求があいまいであった場合にも検索できるようにしている。</p> <p>実験により、実際に高い精度の数式検索が可能であることが示された。しかし、数学的推論を一切していないことや、高度なパターンマッチが未実装であることから、検索不可能な式も存在するなど、実用的なシステムの実現にあたっては、多くの課題が残っていることも確認された。</p>			