

Jozef Mišutka: Mathematical Search Engine

Autor zpracovává novou oblast vyhledávacích strojů – vyhledávání v matematickém textu. Specifičnost tohoto oboru je dána především neexistencí univerzální kanonické podoby matematického výrazu. Práce obsahuje návrh vyhledávacího systému EgoMath, ověření jeho schopností na větší datové kolekci matematických textů, a v neposlední řadě i srovnání s dalším matematickým vyhledávačem.

Samotný výzkum byt rozdělen do tří navazujících oblastí: popis stávajících technik pro vyhledávání v matematice (str. 3-19), návrh nové efektivní techniky (str. 21-84) a srovnání s dalšími existujícími technikami (str. 72-81, 85-100).

Práce obsahuje zajímavý historický přehled vývoje technologií pro vyhledávání v matematickém textu (str. 23-35) a samotných vyhledávacích systémů (str. 39-41). Ukazuje se na něm, že se jedná o poměrně novou disciplínu s mnoha rozdílnými přístupy k řešení.

Zvolený postup v EgoMath je založen na převodu matematických výrazů na sekvence tokenů, které lze již snadno indexovat stávajícími technikami full-textových vyhledávačů. Při vhodných tokenech je tak možné pomocí Boolského modelu najít potencionální kandidáty a podrobit je detailnějšímu zkoumání shody, resp. stanovení hodnoty podobnosti. Stejně tak je možné rozšířit původní matematický zápis o všechny ekvivalentní formy a indexovat pak všechny varianty. Jádrem ale i tak zůstává samotný převod matematického výrazu na sekvenci tokenů a nové postupy a přístupy ke generování příslušných ekvivalentních forem zápisu. Tyto postupy a přístupy jsou popsány a vyhodnoceny v hlavní části práce (str. 21-84).

Z důvodů efektního srovnání s EgoMath, byla v poslední části reimplementována technika Ma et al. s některými novými vylepšeními. Srovnání ukázalo výhodnost systému EgoMath a použitých algoritmů pro vyhledávání v matematickém textu.

Práce je napsána v kvalitní angličtině a má 164 stran včetně příloh. Její výstupy byly prezentovány na několika konferencích a příslušné články jsou rovněž citovány v pracích výzkumníků, kteří působí v téže vědní oblasti.

Z těchto důvodů doporučuji, aby byla práce uznána jako práce disertační.

RNDr. Leo Galamboš, Ph.D.

Ve Svobodě nad Úpou, dne 28.7.2013