

**Mgr. Jiří Mírovský:**  
**Netgraph – a Tool for Searching in the Prague Dependency**  
**Treebank 2.0**  
**Disertační práce**

(vyjádření školitele)

Předložená disertační práce se zabývá vytvořením vhodného vyhledávacího modelu pro prohledávání komplexně anotovaných (závislostních) stormů reprezentujících anlyzu vět přirozeného jazyka. Tento model vychází z požadavků specifikovaných uživateli, především z oblasti teoretického lingvistického výzkumu. Je třeba zdůraznit, že se jedná o problematiku komplexní, a to jak z teoretického hlediska (co je vůbec možno efektivně vyhledávat), tak z hlediska praktického (algoritmy a implementace). Ve světovém měřítku se obdobnou problematikou zabývá jen několik pracovišť, a to zejména Linguistic Data Consortium a University of Melbourne (prof. S. Bird). Autor prací podstatně rozšiřuje systém Netgraph, který vyvinul dříve pro dotazování nad jednodušší reprezentací použitou na analytické rovině Pražského závislostního korpusu.

Práce je rozdělena do devíti kapitol. V úvodu je popsáno zadání práce a krátký popis reprezentace jazykového významu na tektogramatické rovině, jako nejkompexnější lingvistická anotace, na níž byl dotazovací systém konstruován. Ve druhé kapitole autor rozebírá existující vyhledávací systémy a specifikuje tzv. lingvistické požadavky na vyhledávací systém. Třetí kapitola je vlastním jádrem práce: v ní autor odpovídá na zadané požadavky tím, že specifikuje dotazovací jazyk, ve kterém je možno se ptát na všechny lingvistické jevy z původního zadání. Čtvrtá kapitola popisuje použitá data a jejich konverze do formátu, který je využitelný v systému Netgraph. Následující tři kapitoly popisují vytvořený model a rovněž jeho implementaci (systém Netgraph a jeho Netgraph Query Language). V osmé kapitole autor popisuje zkušenosti uživatelů Netgraphu – systém je již v širokém používání a řada dodatečných požadavků přicházela od jeho uživatelů. V poslední kapitole autor shrnuje dosažené výsledky. Další části práce pak obsahují seznam literatury a šest příloh s technickou specifikací dotazovacího jazyka Netgraph, formáty dat, a instrukce pro instalaci z příloženého CD.

Hodnocení:

Práce je kombinací teoretické práce ve vyhledávání v komplexních grafických strukturách, lingvistické analýzy ve vztahu k dané reprezentaci a rovněž aplikací „user studies“ na problematiku vyhledávání (včetně vytvoření vhodného dotazovacího jazyka a interaktivního uživatelského prostředí). V tomto smyslu je to tedy práce interdisciplinární.

Z teoretického hlediska práce přináší srovnání síly vyhledávacích algoritmů z předchozích prací s vytvořeným modelem a algoritmy použitými v Netgraphu. Netgraph integruje všechny dosud známé algoritmy a možnosti vyhledávání, a to i velmi obtížných konstrukcí (tranzitivní závislost, počty závislých uzlů, odkazování mezi uzly). V tomto směru jde tedy o podstatný krok dopředu.

Z uživatelského hlediska je třeba ocenit, že autor spolupracoval velmi úzce s uživateli Netgraphu a dostával tak řadu podnětů a rovněž dostával zpětnou vazbu jak pro vlastní algoritmy vyhledávání, tak i pro uživatelské interaktivní rozhraní. Lze konstatovat, že i zde úkol splnil – řada lingvistických prací vznikla za velmi vydatného přispění tohoto nástroje.

Je třeba rovněž připomenout, že podstatná část práce (její jádro) byla již publikována i na hlavní světové konferenci oboru Computational Linguistics (Columbus, Ohio, USA), a to v jedné z hlavních sekcí.

Práce je psána velmi dobrou angličtinou, a snadno se čte. Obsahuje řadu příkladů a obrázků, které přispívají k pochopení i těch složitějších dotazů a jejich vyhodnocení.

Práce je formálně zcela v pořádku, seznam citované literatury je dostatečný.

Předloženou disertaci Mgr. Jiřího Mírovského tedy považuji za práci, která splňuje nároky kladené na disertační práci a doporučuji ji k obhajobě.

Praha, 22.7.2008

