



UPPSALA  
UNIVERSITET

## Bioinformatics Engineering Program

Uppsala University School of Engineering

<b>UPTEC X 10 022</b>		<b>Date of issue 2010-10</b>
Author <b>Gunnar Dahlberg</b>		
Title (English) <b>Implementation and evaluation of a text extraction tool for adverse drug reaction information</b>		
Title (Swedish)		
Abstract <p>A text extraction tool was implemented on the .NET platform with functionality for preprocessing text (<i>removal of stop words, Porter stemming and use of synonyms</i>) and matching medical terms using permutations of words and spelling variations (<i>Soundex, Levenshtein distance and Longest common subsequence distance</i>). Its performance was evaluated on both manually extracted medical terms (semi-structured texts) from summary of product characteristics (SPC) texts and unstructured adverse effects texts from Martindale (i.e. a medical reference for information about drugs and medicines) using the WHO-ART and MedDRA medical term dictionaries. Results show that sophisticated text extraction can considerably improve the identification of ADR information from adverse effects texts compared to a verbatim extraction.</p>		
Keywords <p>Text extraction, Adverse drug reaction, Permutation, Soundex, Levenshtein distance, Longest common subsequence distance, Porter stemming</p>		
Supervisors <b>Tomas Bergvall</b> Uppsala Monitoring Centre <b>Niklas Norén</b> Uppsala Monitoring Centre		
Scientific reviewer <b>Mats Gustafsson</b> Uppsala University		
Project name	Sponsors	
Language <b>English</b>	Security	
<b>ISSN 1401-2138</b>	Classification	
Supplementary bibliographical information	Pages <b>66</b>	
<b>Biology Education Centre</b> Box 592 S-75124 Uppsala	<b>Biomedical Center</b> Tel +46 (0)18 4710000	<b>Husargatan 3 Uppsala</b> Fax +46 (0)18 471 4687