ScholarWorks

An exploratory study using the predicate-argument structure to develop methodology for measuring semantic similarity of radiology sentences

Newsom, Eric Tyner

Advisor:	Jones, Josette F.
	Gamache, Roland E.
	Mahoui, Malika
Degree:	M.S.
Degree Year:	2013
Department:	School of Informatics
Grantor:	Indiana University
Permanent Link:	http://hdl.handle.net/1805/3666

Keywords:

Natural Language Processing ; Information Extraction ; Predicate-Argument Structure ; Semantic Similarity

LC Subjects:

Computational linguistics -- Analysis ; Natural language processing (Computer science) ; Semantic computing -- Research ; Semantics -- Data processing ; Description logics ; Electronic information resource searching -- Research ; Data mining ; Semantic Web ; Text processing (Computer science) ; Information storage and retrieval systems -- Research ; Predicate (Logic) ; Medical informatics -- Data processing

Date: 2013-11-12

Abstract:

The amount of information produced in the form of electronic free text in healthcare is increasing to levels incapable of being processed by humans for advancement of his/her professional practice. Information extraction (IE) is a sub-field of natural language processing with the goal of data reduction of unstructured free text. Pertinent to IE is an annotated corpus that frames how IE methods should create a logical expression necessary for processing meaning of text. Most annotation approaches seek to maximize meaning and knowledge

by chunking sentences into phrases and mapping these phrases to a knowledge source to create a logical expression. However, these studies consistently have problems addressing semantics and none have addressed the issue of semantic similarity (or synonymy) to achieve data reduction. To achieve data reduction, a successful methodology for data reduction is dependent on a framework that can represent currently popular phrasal methods of IE but also fully represent the sentence. This study explores and reports on the benefits, problems, and requirements to using the predicate-argument statement (PAS) as the framework. A convenient sample from a prior study with ten synsets of 100 unique sentences from radiology reports deemed by domain experts to mean the same thing will be the text from which PAS structures are formed.

Description:

Indiana University-Purdue University Indianapolis (IUPUI)

Files in this item

Name: Thesis_Eric_Newsom.pdf
Size: 3.513Mb
Format: PDF
Description: Main thesis document

[View/Open](#)

This item appears in the following Collection(s)

- Informatics Graduate Theses and PhD Dissertations

Fulfilling the Promise

Privacy Notice