

Applying Rich Ontology to QA & IR

Su Nam Kim

CSSE / University of Melbourne

July 19-20, 2007, Sydney

HCSNet Workshop

on Next Generation Search Technology



HCSNet

The ARC Network in Human Communication Science

Research Background (I)

- **Ph.D Research Topic**
 - **Statistically Modeling MWEs**
- **Brief Information of MWEs**
 - **Def)** lexical items that can be decomposed into multiple simplex words
 - **Types in English MWEs**
 - Compound Nouns(*information retrieval*)
 - Verb-Particle Constructions(*take off*)
 - Light-Verb Constructions(*take a walk*)
 - non-decomposable idioms(*kick the bucket*)

Research Background (II)

- **Computational Tasks for MWEs**
 - Identification/Extraction of MWEs
 - Degree of Compositionality of MWEs
 - Interpretation of Semantics of MWEs
 - crossover/crosslingual study
- **Approaches**
 - Statistically modeling computational tasks of MWEs using Substitutability, Distributional & Semantic Similarity, Co-occurrence, Interpretationality and Linguistic features

Background

- **Focused Issues for QA, IR w.r.t MWEs**
 - MWEs as the rich ontology to provide clues for QA
 - MWEs as the enriched queries
 - MWEs to rank relevant documents
- **Aim to use MWEs in QA, IR**
 - provide rich queries based on MWE compositionality
 - provide the rich ontology to understand questions or question style queries
 - provide the clues to answer

Motivation (I)

- Applying Rich Ontology for QA and IR which takes question type queries
- Understand Syntax and Semantics of Units in domain specific Search Engine (e.g. medical domain)
 - based on **Semantic Relations(SRs)** of compound nouns, extract the clue to answer
 - e.g. *What is pie made of?*
 - “*apple pie*” : pie made of apple
 - what = apple

Background of SRs (I)

- **Def) Semantic Relations (SRs)**
 - express the relation between the head noun and its modifier(s)
- **Examples of SRs**
 - *orange juice* (**MAKE**), *student loan*(**BENEFIT**), *fair penguin*(**PROPERTY**), *morning class*(**TIME**)
 - Levi (1979), Finin (1980), Vanderwended (1994), Barker (1998), Rosario (2002) etc.
 - pragmatism : *She put the skirt into cotton bag*

Background of SRs (II)

- **Tasks related SRs**

- Syntactic disambiguation(bracketing)

- e.g. *((computer science) department) vs (linguistics graduate program)*

- Seek for a set of SRs

- Interpretating compound nouns(CNs)

- e.g. *orange juice(MAKE) vs morning juice(TIME)*

- Word Sense Disambiguation for CNs

- e.g. *art museum : WS1*

Applications : QA (I)

- Provide clues to answer
- Extract the semantic units
- *e.g.1 Where have nuclear incidents occurred?*
- key : *where* indicates **LOCATION**
- answer : *Three Mile Island* of the sentence, “The Three Mile Island nuclear incident caused a DOE policy crisis.”

Applications : QA (II)

- *e.g.2 What did the factory in Howell Michigan make?*
- key: *what*= an element in noun phrase, SR= **MAKE**
- answer : **Car** of the sentence, “The car factory in Howell Michigan closed on Dec. 22, 1991.”

Applications : QA (III)

- **Combine it with particular type of QA**
- **WH-questioned QA for news media (e.g. *News In Essence* <http://www.newsinesence.com>)**
- WH-question based SRs (Vanderwende 1994)
- **e.g. *Where were the drown victims found?***
- key : **where** indicates Location
- Answer : the **bay** of the sentence, “The bodies of the three men and a woman were pulled from the bay.”

Applications : Biomedical Search Engine (I)

- **Interpreting SRs in CNs in bio domain**
- **Understand the semantic units and relations in units** (Rosario 2002, Grover 2005, Nakov 2006)
- e.g. noun phrases with SRs
 - **Cause** : *asthma hospitality, automobile accident*
 - **Defect** : *hormone deficiency, blood plasma*
 - **Location** : *brain artery, emergency surgery*
- **e.g. query for “migraine treatment”**
 - pages containing good paraphrasing verbs like “**relieve**” or “**prevent**” would be preferred (treatment that relieve migraine)

Applications : Biomedical Search Engine (II)

- **Syntactic disambiguation for NCs (Nakov 2005)**
- **Understand the semantic units & Provide units for interpreting NCs**
- e.g. examples of bracketing
 - Left Bracketing : *((liver cell) antibody)*
 - Right Bracketing : *(liver (cell line))*

Implications

- Handling SRs in CNs provides syntactically and semantically rich ontologies for QA & IR
- Provide Syntactic and Semantics information of Units for Domain Specific Search Engine (e.g. biomedical domain)
- Improve Accuracy and Fluency of QA & IR