Internet Search: Semantics vs. "Semantics"

Christian F. Hempelmann

Victor Raskin

Human Language Semantics

- **Information comes as Natural Language (NL)**

- **No Search Relevance without Understanding NL**

  - Underlying production and comprehension rules

  - Users with low error tolerance

  - Observable output in principle irregular

- **No Understanding NL without Semantics**

hakia®

Processing NL Done "Semantically"

▶ **Convert NL into FOL (e.g., "Logic Forms")**

▶ **Count co-occurence (e.g., LSA)**

▶ **Characteristically, this lack of interest in linguistic theory expresses itself in the proposals to limit the term 'theory' to 'summary of data' [...]
(Chomsky 1965: 194)**

▶ **Not all that is measurable is meaning!
(Lyons 1963: 5)**

hakia

Done Semantically: OntoSem

- **7k-concept language-independent ontology**

- **Ontology-based lexicons, including a 50k-entry English lexicon with 80k senses**

- **Onomastica, dictionaries of proper names**

- **Text meaning representation (TMR) language, an ontology-based knowledge representation language,**

- **OntoParser transforming NL text**

hakia

Semantics is Expensive: Offline Processing

▸ **OntoParse crawled pages**

▸ **Understand their meaning**

▸ **Anticipate queries about this meaning**

```
outlaws ran cocaine into united-states.

ran : smuggle : 3.5
        agent           : outlaws : criminal        : 2 : 8
        instrument      : cocaine : cocaine          : 1 : 5
        location        : united-states : country    : 1 : 4


ran : operate-device : 3
        agent           : outlaws : criminal        : 1 : 3
        theme           : united-states : country    : 1 : 4
        instrument      : cocaine : cocaine          : 1 : 5

ran : run-for-office : 2.5
        agent           : outlaws : criminal        : 2 : 3
        theme           : united-states : country    : 1 : 3
```

▐ Where was the cocaine smuggled?

▐ What was shipped to the United States?

hakia

BETA

hakia®

making sense of the World's information