

Closing the Semantic Gap

Marc Vilain

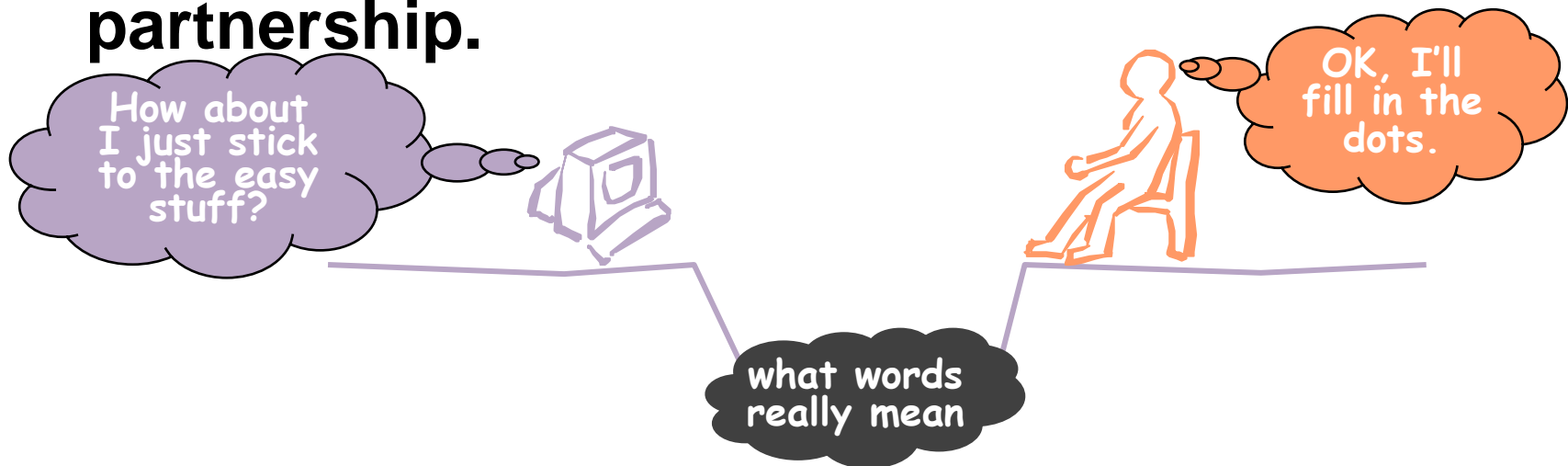
781-271-2151 • mbv@mitre.org

MITRE Sponsored Research



Problem: the Semantic Gap

- Language-enabled applications are limited by their ability to understand the actual meanings of words.
- That's why analysis is a human-machine partnership.



- In mission-critical analytic situations, there is often not enough human power to fill in all the dots, all the time, for all the relevant data.

MITRE

Background: Events Are Key

The **United States** **Friday** condemned the bomb attack on a **Damascus** bus this week, after the **Syrian government** lashed out at it for failing to do so and for dismissing **Syrian** accusations that **Israel** was behind the bombing.

Syria
United States
Israel ...



Johnny Name-tagger

Terrorism,
diplomacy,
or both



Tommy Topic-tracker

Bombing
Agent=Syria



Herbie Heuristic-extractor

Current methods only give a *partial* view of the events in a text: we need new ways of detecting and understanding events.

MITRE

© 2007, The MITRE Corporation

Objective

- **Key theme: process the meaning of events**
 - Identify events
 - Find what's new about an event
 - Find the outcomes of events
- **Approach: focus on word meanings**
- **Method: create explicit semantic database**
 - Linguistically – machine-readable dictionaries
 - Logically – ontologies, hand-built models
 - Statistically – real-life distributions

Activities

- **Lexibase: semantic knowledge base built from multiple dictionaries and ontologies**
- **Dictionary-speak: processing the language of dictionary glosses**
- **Event matching: statistical exploitation of Lexibase**
- **Rummage: event search and extraction demo**
- **Evaluation: exploit standard information retrieval evaluations to assess Lexibase value**

Highlight: Semantic Matching

The United States Friday condemned the **bomb attack** on a **Damascus bus** this week, after the Syrian government lashed out at it for failing to do so and for dismissing Syrian accusations that Israel was behind the bombing.

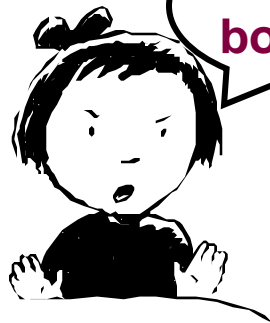
Damascus: the capital of **Syria**

Syria: a country..., adj. **Syrian**



New Oxford dictionary

bomb, vb: to **attack** with a **bomb**,
(noun: **bombing**)



**Syrian
bombing**

MITRE

© 2007, The MITRE Corporation

Demonstration: Event search

Rummage syrian bombings

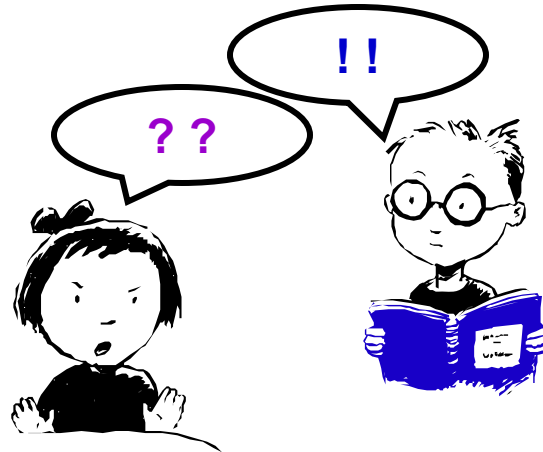
Found 50 documents (28041 msec)

/XIE19970118.0109.xml score:0.8763088	XIE19970104.0183 1997-01-04 U.S. Condemns Bus Bombing in Damascus
/XIE19970219.0112.xml score:0.7916559	WASHINGTON, January 3 (Xinhua) -- The United States on Friday condemned the bomb attack on a Damascus bus this week after the Syrian government lashed out at it for failing to do so and for dismissing Syrian accusations that Israel was behind the bombing.
/XIE19970104.0183.xml score:0.78664005	"The United States condemns this terrorist attack against innocent people, and there's every reason to believe that these were innocent people on that bus, civilians, who ought not to have to fear for their lives when they travel around the city of Damascus," State Department spokesman Nicholas Burns said at a news briefing.
/XIE19970103.0267.xml score:0.7595911	Burns comment came hours after the Syrian Foreign Ministry issued a statement blasting him for his reaction to the bomb attack on the bus, that killed nine people and injured 44.
/XIE19971008.0053.xml score:0.7377309	At thursday's briefing, Burns, without expressing any condolences to the Syrian government for the people killed in the attack, warned Syria against accusations of Israeli agents for the bomb attack, saying: "It is a very serious charge for one state to accuse another of actually sponsoring bombings in its capital... it's a wild and irresponsible charge if in fact it is not backed up by evidence."
/XIE19970225.0325.xml score:0.72789973	In reaction to burns' Thursday remarks, the Syrian statement said: "We noted the absence of any condemnation for this terrorist attack in the statement of the spokesman Burns, and of any human sympathy with the families of the victims.
/XIE19970118.0240.xml score:0.7216116	
/APW19981021.1162.xml score:0.659993	
/XIE19970102.0032.xml score:0.65679526	
/XIE19960703.0156.xml score:0.65576077	
/XIE19960811.0055.xml score:0.6555333	
/XIE19970104.0084.xml score:0.6555333	
/APW19980824.0892.xml score:0.6420001	
/XIE19991210.0039.xml score:0.6409544	
/XIE20000114.0040.xml score:0.6312022	
/XIE19970102.0041.xml score:0.6222271	
/XIE19980704.0022.xml score:0.6076729	
/APW19980703.0410.xml score:0.59629095	
/APW19980703.0544.xml score:0.59629095	
/XIE19970915.0032.xml score:0.59629095	
/XIE19970915.0112.xml score:0.59629095	
/XIE19960307.0315.xml score:0.5782202	
/XIE19970305.0178.xml score:0.57379067	

Impacts

- **Technical: finding/extracting events in the large, with greater depth, more smarts, and less training**
- **New capabilities: What's new about events? What's the outcome of events?**
- **Applications to**
 - **Indications and warnings**
 - **Summarization, chronology creation, the Atlas of Grievances**
 - **Indexing and curation of text databases**

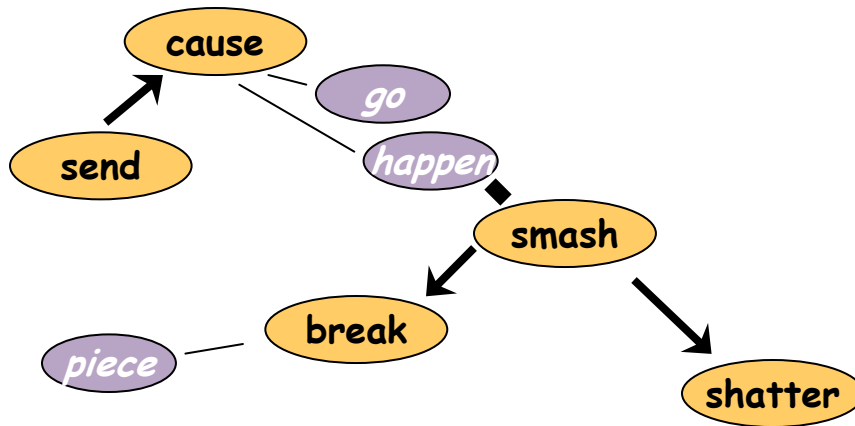
Future Plans



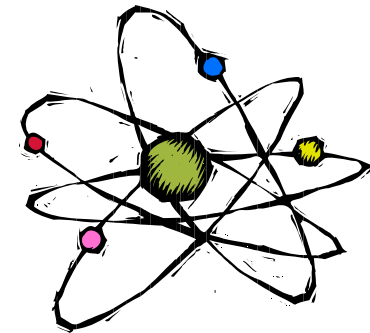
Semantic search experiments



Evaluation against standard test suites



Word association models based on human memory



Ontologies of primitive semantic elements

MITRE

© 2007, The MITRE Corporation