Information Extraction and Name Entity Recognition

By Sean Humpherys
In the future, everything will be effortless!

Computers will take care of every task. We'll just point to what we want done and click. We'll never need to leave the climate-controlled comfort of our homes!

No nuisance, no wasted time, no annoying human interaction...

...no life. Life is too inconvenient.
Information Extraction

- Name entity recognition
- Reference resolution
- Relation detection and classification
- Event detection and classification
- Temporal expression detection
Examples

- These results suggest that con A-induced [DISEASE hepatitis] was ameliorated by pretreatment with [TREATMENT TJ-135].

- [DISEASE Malignant mesodermal mixed tumor of the uterus] following [TREATMENT irradiation] shrank...
Information Extraction

- Name entity recognition
- Reference resolution
- Relation detection and classification
- Event detection and classification
- Temporal expression detection
“United Airlines said Friday it has increased fares by $6 “

Who?
- United Airlines
- Two words with initial letter capitalized
NER Strategies

- Presence in a name entity lists
- Gazetters, millions of location entities, geographical, geological and political.
  - http://www.geonames.org

- United States Census Bureau list of first names and surnames
  - http://www.census.gov/genealogy/www/freqnames2k.html

- Titles: Rev. MD, Inc.
## Strategy using Shapes

<table>
<thead>
<tr>
<th>Capitalization</th>
<th>Tim Wagner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>American Airlines</td>
</tr>
<tr>
<td>All Caps</td>
<td>IRA</td>
</tr>
<tr>
<td>Mixed Case</td>
<td>eBay</td>
</tr>
<tr>
<td>Capitalized initial with period</td>
<td>Inc., Corp., S.A. de C.V</td>
</tr>
<tr>
<td>Ends in digits</td>
<td>T3</td>
</tr>
</tbody>
</table>
**United Airlines** said Friday it has increased fares by $6. American Airlines immediately matched the move. **United**, a unit of UAL Corp., said the increase took effect Thursday.

- United Airlines and United refer to the same entity
United Airlines said Friday it has increased fares by $6. American Airlines immediately matched the move. United, a unit of UAL Corp., said the increase took effect Thursday.

United Airlines is related to UAL Corp.
United Airlines said Friday it has increased fares by $6. American Airlines immediately matched the move. United, a unit of UAL Corp., said the increase took effect Thursday.
Temporal Detection

United Airlines said Friday it has increased fares by $6. American Airlines immediately matched the move. United, a unit of UAL Corp., said the increase took effect Thursday.
Fare-Raise Attempt:
[ Lead Airline: United Airlines
Amount: $6
Effective Date: 2006-10-26
Follower: American Airlines ]