SpatialML: Annotation Scheme, Corpora, and Tools

Inderjeet Mani, Janet Hitzeman

Justin Richer, Dave Harris

Rob Quimby, Ben Wellner

The MITRE Corporation

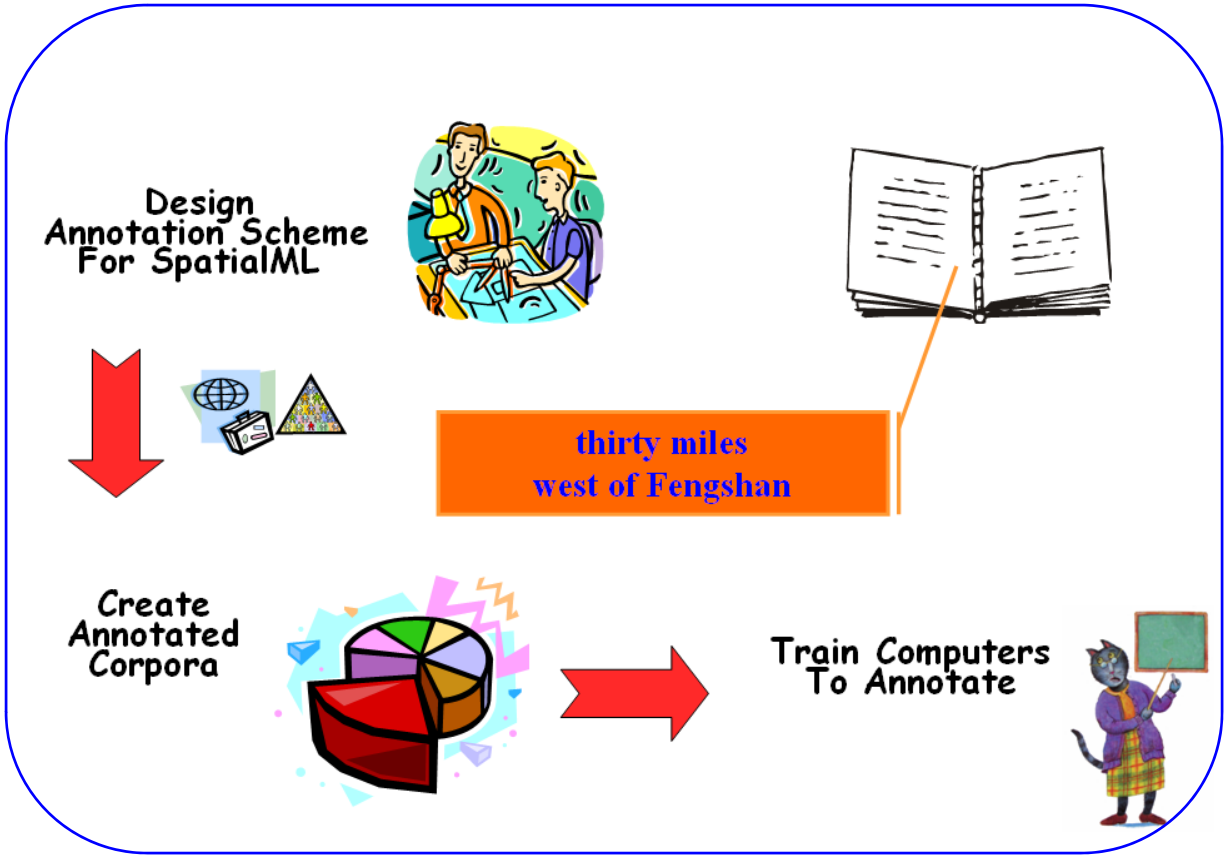
imani at mitre.org

Motivation

- The interpretation of spatial language has been hampered by the lack of a markup scheme
- As a result, lack of resources such as corpora and evaluation methods for systems that process spatial language
- SpatialML is a markup scheme for representing places mentioned in text and their relationships
- The main focus has been on geo-coding of natural language, i.e., the mapping of geographic references in text to data in gazetteers and other databases.

sourceforge.net/projects/spatialml

Some Challenges in Annotation-Based Methods



1. ***Annotation Scheme: Expressiveness versus Usability***
2. ***Maturity of Guidelines***
3. ***System Adaptation Cost***
   - ***Languages***
   - ***Domains***

SpatialML Example

a <PLACE id="1" type="FAC" form="NOM">building</PLACE>

<SIGNAL id="2">5 miles</SIGNAL>

<SIGNAL id="3">east</SIGNAL>

of <PLACE id="4" type="PPL" country="TW" form="NAM"
latLong="22°37'N 120° 21'E">Fengshan</PLACE>

<PATH id="5" source="4" destination="1" distance="5:mi"
direction="E" signals="2 3" frame="EXTRINSIC"/>

Multilingual Examples

I live in a [town] some [50 miles] [south] of [Salzburg] in the central [Austrian] [Alps].

جبال الالب النمسا و سالزبرج فى وسط جنوب خمسين ميل مدينة تبعد حوالى أنا أسكن فى

<PLACE type="PPL" id=1 form="NOM">مدينة</PLACE>

<SIGNAL id=2>خمسين ميل </SIGNAL>

<SIGNAL id=3>جنوب /SIGNAL>

<PLACE id=4 type="PPLA" country="AT" form="NAM">سالزبرج </PLACE>

<PLACE id=5 type="COUNTRY" country="AT" mod=C>النمسا <PLACE>

<PLACE id =6 type="MTS">جبال الالب </PLACE>

<PATH id=7 distance="50:mi" direction=S source= 4 destination=1 signals="2 3"/>

<LINK id=8 source=1 target=6 linkType="IN"/>

나는 **[오스트리아] [알프스]** 중심의 **[잘츠부르크] [남쪽]**에서 **[50**마일**]** 거리의 마을에 산다

<PLACE type="PPL" id=1 form="NOM" ctv="TOWN">마을**</PLACE>**

<SIGNAL id=2>50 마일**</SIGNAL>**

<SIGNAL id=3>남쪽**</SIGNAL>**

<PLACE id=4 type="PPLA" country="AT" form="NAM">잘츠부르크**</PLACE>**

<PLACE id=5 type="COUNTRY" country="AT" mod="C">오스트리아**</PLACE>**

<PLACE id=6 type="MTS" >알프스**</PLACE>**

<PATH id=7 distance="50:mi" direction=S source= 4 destination=1 signals="2 3"/>

<LINK id=8 source=1 target=6 linkType="IN"/>

MITRE

PLACE TYPES

- Coarse-grained, to make it easier for humans and machines to annotate
- Drawn opportunistically from
  - Alexandria Digital Library Feature Types Thesaurus
  - NGA Geonames
  - USGS GNIS

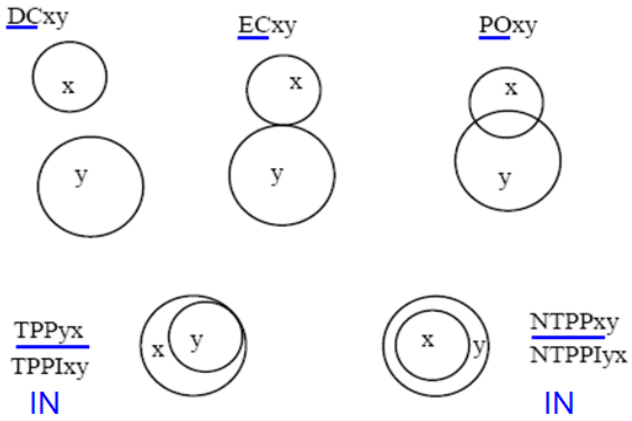
BODYOFWATER	River, stream, ocean, sea, lake, canal, aqueduct, geyser, etc.
CELESTIAL	sun, moon, Jupiter, Gemini, etc.
CIVIL	Political Region or Administrative Area, usually sub-national, e.g. State, Province, certain instances of towns and cities.
CONTINENT	Denotes a continent, including ancient ones. See Table 2.
COUNTRY	Denotes a country, including ancient ones. See Table 1.
FAC	Facility, usually a catchall category for restaurants, churches, schools, ice-cream parlors, bowling alleys, you name it!
GRID	A grid reference indication of the location, e.g., MGRS (Military Grid Reference System)
LATLONG	A latitude/longitude indication of the location
MTN	Mountain
MTS	Range of mountains
POSTALCODE	Zipcodes, postcodes, pincodes etc.
POSTBOX	P. O. Box segments of addresses
PPL	Populated Place (usually conceived of as a point), other than PPLA or PPLC
PPLA	Capital of a first-order administrative division, e.g., a state capital
PPLC	Capital of a country
RGN	Region other than Political/Administrative Region
ROAD	Street, road, highway, etc.
STATE	A first-order administrative division within a country, e.g., state, province, gubernia, territory, etc.
UTM	A Universal Transverse Mercator (UTM) format indication of the location
VEHICLE	Car, truck, train, etc.

Relations between PLACEs

a <PLACE id="1" form="NOM" type="FAC">school</PLACE> in
<PLACE id="2" form="NAM" type="PPL" latLong="39.952°N 75.164°W">Philadelphia</PLACE>

<LINK source=1 target=2 linkType="IN"/>

RCC8 Spatial Calculus*



*http://www.irit.fr/~Philippe.Muller/Publis/ci02.pdf

LinkType	Example
IN (tangential and non-tangential proper parts)	[Paris], [Texas]
EC (extended connection)	the border between [Lebanon] and [Israel]
NR (near)	visited [Belmont], near [San Mateo]
DC (discrete connection)	the [well] outside the [house]
PO (partial overlap)	[Russia] and [Asia]
EQ (equality)	[Rochester] and [382044N 0874941W]

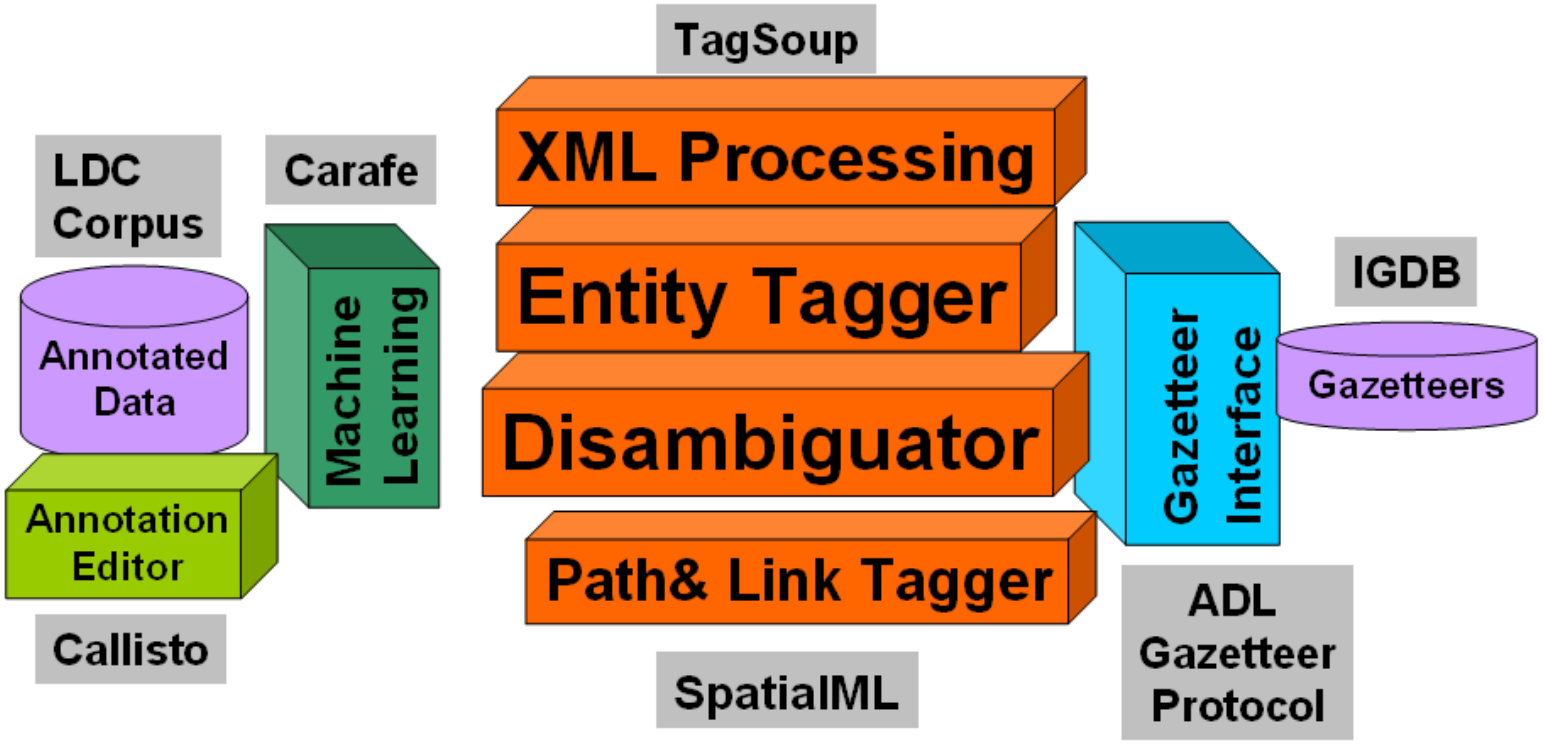
Orientations

MOD Code	Example
B	the bottom of the [well]
BR	[Burmese] border
C	central [district]
E	eastern [province]
N	[North India]
NEAR	near [Harvard]
S	southern [India]
T	the top of the [mountain]
W	west [Tikrit]

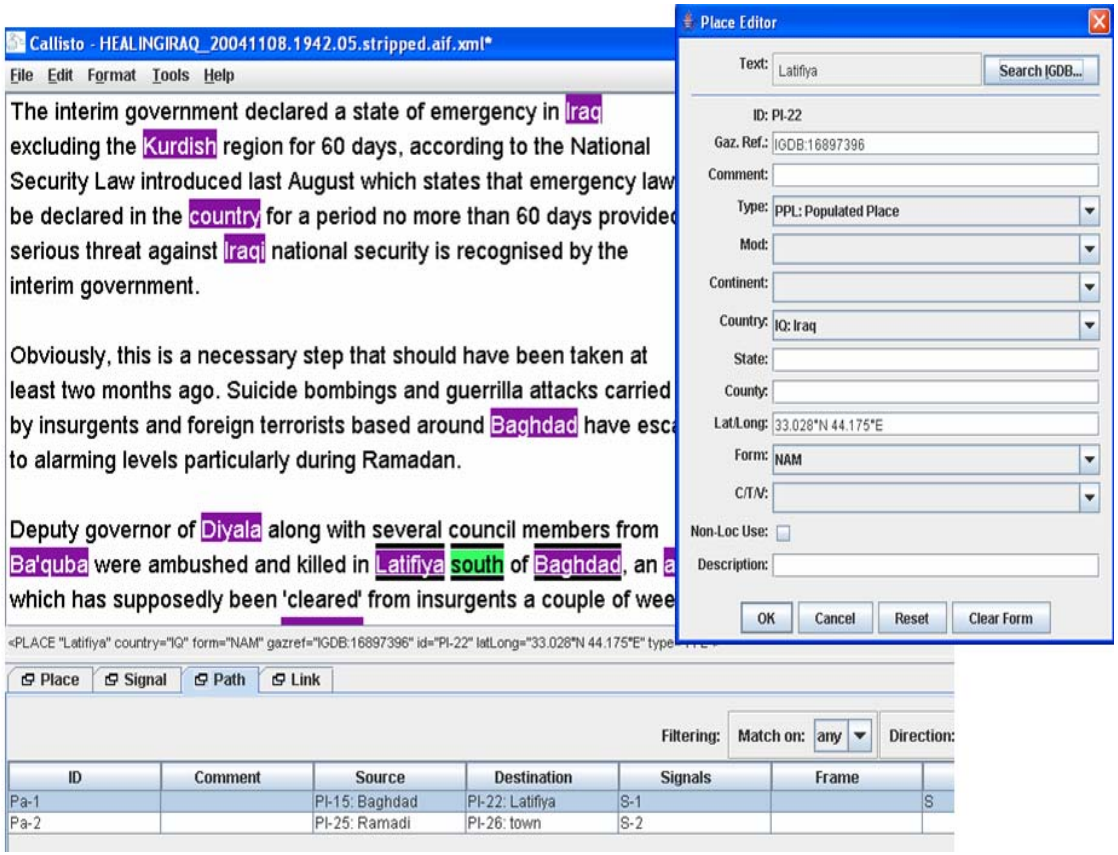
Direction Code	Example
B	[behind] the house
A	[above] the roof
BL	[below] the tree-line
E	[E] of
ESE, WSW, etc.	
F	[in front of] the theater
N	[north] of
S	[south] of
W	[W] of

MITRE

MIPLACE Open Architecture



Annotation Environment and Corpora



callisto.mitre.org

- ***ACE ASC -- 428 docs***
  - **LDC2008T03**
- ***ProMED emerging diseases-- 100 docs***
- ***US Immigration and Customs (ICE) -- 121 docs***

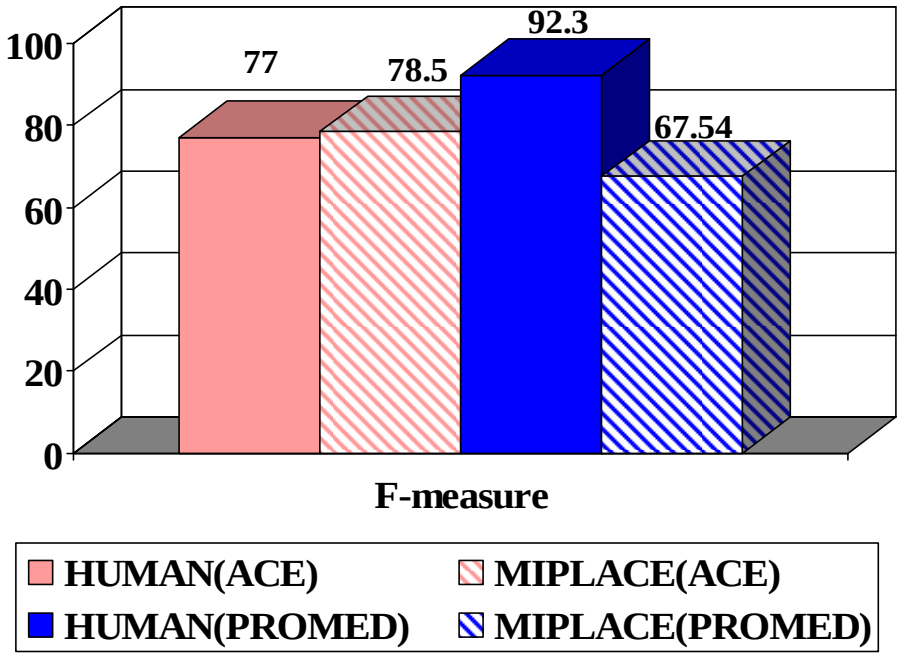
Inter-Annotator Agreement

Attribute	P	R	F
Extent	89.32	95.4	92.3
Form	100	99.14	99.56
LatLong	96.51	57.22	71.85
Gazref	70.44	57.17	63.11

- **Disagreements traced to**
  - **Guidelines**
  - **Expertise**
  - **Gazetteer Interface**
    - **Coverage (not in IGDB, finding via Google)**
    - **Query Language (poor or no support for morphology, transliteration, qualifiers)**

MIPLACE Entity Tagger

a <PLACE id="1" form="NOM">school</PLACE> in <PLACE id="2" form="NAM">Philadelphia</PLACE>



sourceforge.net/projects/carafe

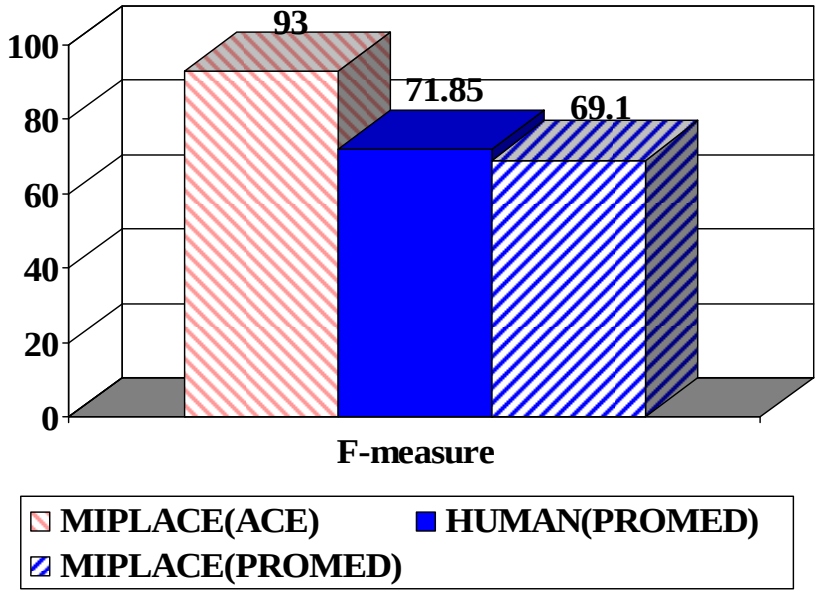
MIPLACE Disambiguator

a <PLACE id="1" form="NOM">school</PLACE> in <PLACE id="2" form="NAM" type="PPL" latLong="39.952°N 75.164°W">Philadelphia</PLACE>

$$\Pr(G_i \mid M) = \frac{e^{\sum_k w_k^* f_k(G_i, M)}}{\sum_{G_j \in Gaz(M)} e^{\sum_k w_k^* f_k(G_j, M)}}$$

$$\arg\max_{G_i \in Gaz(M)} P(G_i \mid M)$$

Log-linear ranker



sourceforge.net/projects/carafe

MIPLACE Path Tagger

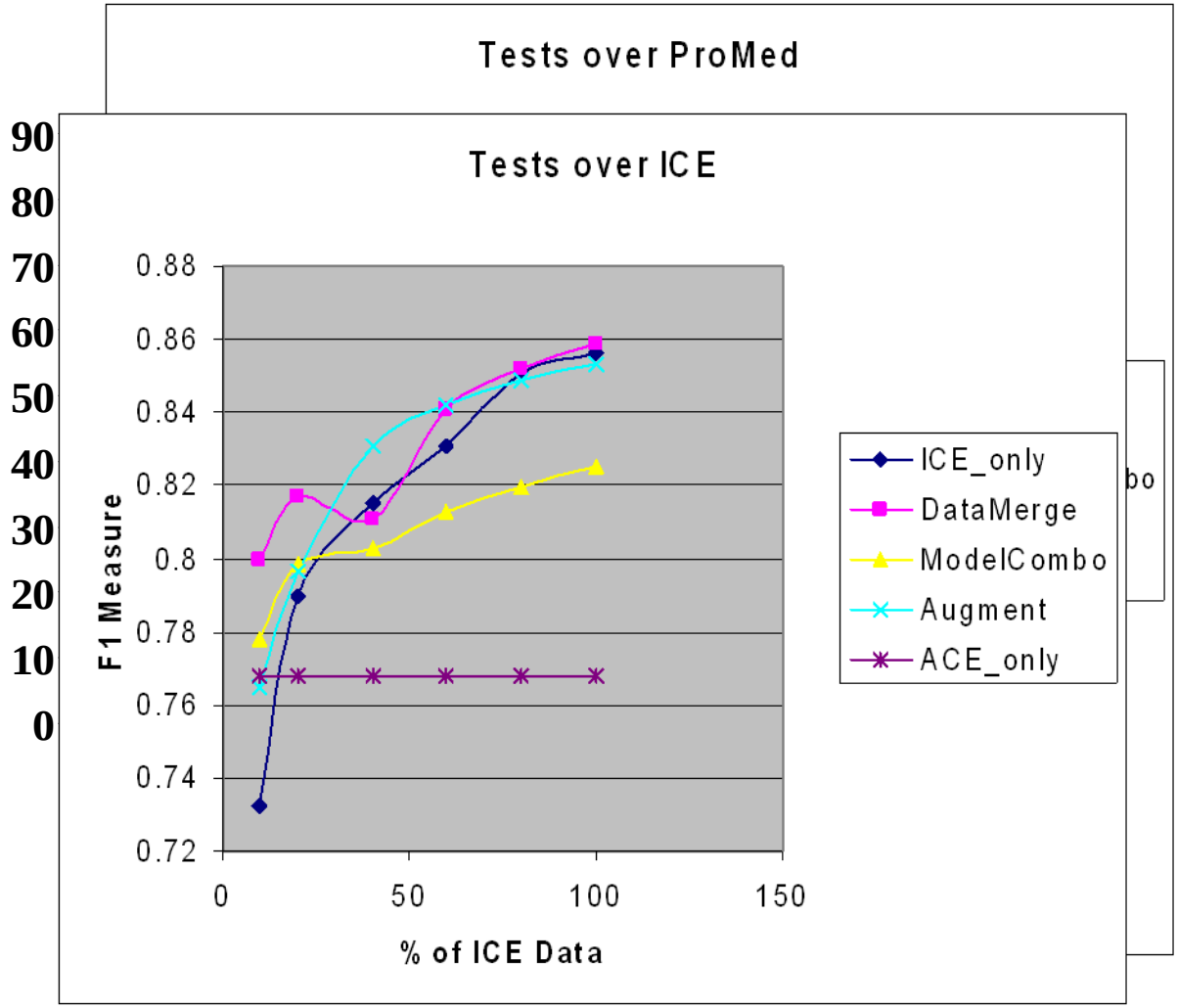
A <PLACE id="1">school</PLACE> <SIGNAL id="2">two hours</SIGNAL> <SIGNAL id="3">north</SIGNAL> of <PLACE id="4">Fengshan</PLACE>

<PATH id="5" source="4" destination="1" distance="2:hr" direction="N" signals="2 3"/>

- Uses rule-based component that recognizes PLACE tags, signals, directions, and distances

MITRE

Domain Adaptation



Related Work

- SpatialML borrows ideas from Schilder et al. (2004) , Garbin and Mani (2005), and Toponym Resolution Markup Language of Leidner (2006).
- SpatialML is compatible with Automatic Content Extraction (ACE) English Annotation Guidelines for Entities (Version 5.6.6 2006.08.01), specifically their GPE, Location, and Facility *entity* tags and the Physical *relation* tags. Unlike ACE, SpatialML:
  - Grounds mentions with geo-coordinates where possible
  - Handles relative locations involving distances and orientation relations
  - Doesn't group mentions into coreference classes
  - Doesn't address metonymy
- SpatialML can be integrated with the Geography Markup Language (GML) defined by the Open Geospatial Consortium (OGC).
- SpatialML leverages ISO (ISO-3166-1 for countries and ISO-3166-2 for provinces).
- Mappings: SpatialML to KML , and from MetaCarta output to SpatialML.

Conclusions

- Developed SpatialML, annotation scheme and guidelines for geo-coding natural language
- Created 3 corpora annotated with SpatialML
- Computed the first large-scale evaluations of guideline-based geo-coding tools
- Evaluated methods for porting across domains
- Future work:
  - MIPLACE Mandarin tagger
  - Integration of SpatialML and TimeML
    - *"Tell me where X has been for the past ten days"*

我居住在一个离中

<PLACE id =1 type="COUNTRY" country="AT" mod="C">奥地利Austrian</PLACE>

<PLACE id =2 type="MTS">阿尔卑斯Alps</PLACE>

<PLACE id=3 type="PPLA" country="AT" form="NAM">萨尔茨堡Salzburg</PLACE>

<SIGNAL id=4>以南south</SIGNAL> 大约 <SIGNAL id=5>50英哩50 miles</SIGNAL> 的

<PLACE type="PPL" id=6 form="NOM" ctv="TOWN">镇子town</PLACE>里。

<PATH id=7 distance="50:mi" direction=S source=3 destination=6 signals="2 3"/>

<LINK id=8 source=1 target=6 linkType="IN"/>

MITRE