

# Local Entity Detection and Recognition Task EVALITA 2009

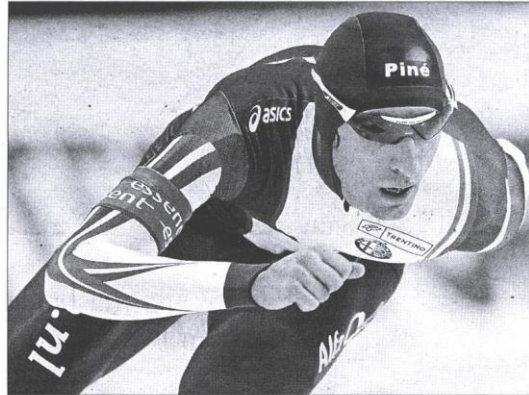
Silvana M. Bernaola Biggio, Claudio Giuliano, Massimo Poesio,  
Yannick Versley, Olga Uryupina, Roberto Zanolì

# Local Entity Detection and Recognition

## Per Enrico Fabris esordio con il record A Salt Lake City sugli scudi anche Chiara Simionato prima nei mille



LA FESTA. Vittoria e record del mondo, c'è di che esultare



POTENTE. Enrico Fabris in azione durante i vittoriosi 5000 a Lake Salt City

di DANIELE FERRARI

TRENTO - Ottimo avvio in coppa del mondo per la nazionale azzurra di pattinaggio velocità, composta anche dai pinetani Matteo Anesi e Ermanno Ioriatti. Due i successi azzurri ottenuti da Enrico Fabris sui 5000 metri sabato e ieri da Chiara Simionato sui 1000 metri, nella prima prova di coppa del mondo disputata nel week-end sul veloce impianto olimpico di Salt Lake City. Con il tempo di 6'07"40 Enrico Fabris ha vinto sabato la gara dei 5000 metri, davanti all'olandese Sven Kramer ed al norvegese Haedse Boeckko, stabilendo inoltre il nuovo

record mondiale della distanza. Nella giornata di venerdì Fabris era giunto invece quinto sui 1500 metri nella gara vinta dall'olandese Wennemars. Positivo anche il terzo posto nella gara dei 1500 metri ottenuto sabato da Chiara Simionato con il tempo di 1'54"65, superata dalle canadesi Christine Nesbitt e Kristina Groves. Nella gara di apertura dei 1000 metri la cadorina Simionato era giunta invece quinta nella prova vinta dalla cinese Wang, e decima sui 500 metri. E' rimasto invece più lontano dal podio il pinetano Matteo Anesi che ha ottenuto il 26° posto sui 1500 metri gruppo A, ed il 29° sui 5000 metri sempre di gruppo A. Gara senza acuti anche per l'altro sprinter

pinetano Ermanno Ioriatti classificatosi al 25° posto sui 500 metri e al 26° sui 1000 metri. Nella giornata conclusiva di ieri dedicata alla prove sprint grande successo di Chiara Simionato sui 1000 metri vinti con il tempo di 1'13"27 davanti alla tedesca Friesinger e alla canadese Nesbitt. Fabris era invece sesto nel gruppo B dei 1000 metri e 14° sui 500 metri di gruppo B, distanze non congeniali all'atleta di Roana. Matteo Anesi chiude la sua fatica nord americana con un doppio 23° posto sui 500 e sui 1000 sempre di gruppo B, anche lui chiamato a confrontarsi sulle distanze sprint non usuali per gli atleti all-round. Nel gruppo A dei 500 metri Ermanno Ioriatti ot-

teneva invece il 22° posto nella prova vinta dal canadese Wotherspoon.

Lo spettacolare circuito di coppa del mondo di pattinaggio velocità vivrà anche quest'anno la sua tappa italiana. Il prossimo 2 e 3 febbraio 2008 sarà l'Ice Rink Piné di Miola di Piné il teatro delle slide mondiali sui 1500 e 5000 metri maschili e femminili con Fabris, Simionato, e si spera i padroni di casa Anesi e Ioriatti, tra i protagonisti più attesi. La prima gara della stagione italiana del pattinaggio si terrà invece il prossimo 24 e 25 novembre sul ghiaccio dell'Ice Rink Piné con la prima prova del «Grand Prix» gara riservata sia agli atleti senior sia agli junior sui 500, 1000, 1500 e 3000 metri.

2 subtasks ....

Entity Mention Detection

- ✓ Detection of the extension of the name
- ✓ Classification

Coreference Resolution

# Entity Mention Detection



- ✓ **Corpus:** I-CAB, 525 news stories from the newspaper L'Adige (07Sep04; 08Sep04; 07Oct04; 08Oct04; Attualita, Cultura, Economia, Sport, Trento)
- ✓ **Classification:** Person, Organization, Location, Geo-Political Entity.
- ✓ **Simple and nested mentions:**
  - ✓ *[La [cantante]NOM\_PER [Madonna]NAM\_PER]NAM\_PER*
  - ✓ *[ [il [presidente]NOM\_PER del consiglio provinciale]NOM\_PER [Giacomo [Bezzi]NAM\_PER] NAM\_PER ]NOM\_PER*

## Training Data

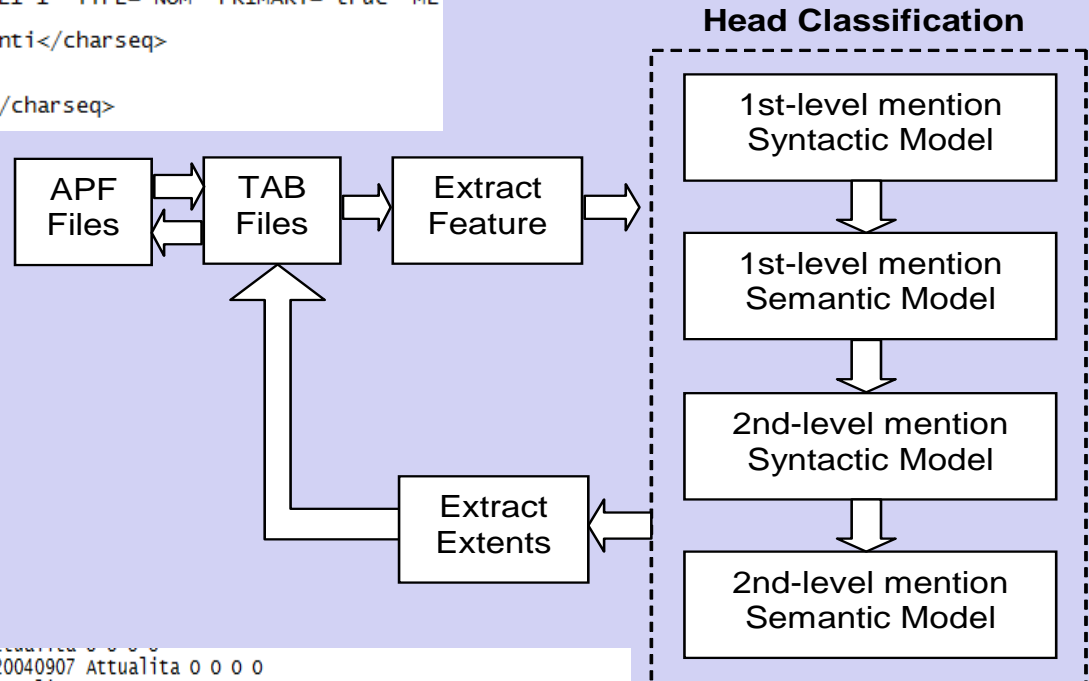
Files	335
Words	113,634
Average words per file	339
Tokens	132,586
<hr/>	
Person	2,938
Organization	2518
Geo-Political	1740
Location	240



# Architecture of the system

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE source_file PUBLIC "SYSTEM" "apf.v5.2.0.dtd">

<source_file URI="adige20040907_id405381.txt" SOURCE="unknown" TYPE="text" VERSION=
<document DOCID="adige20040907_id405381">
  <entity ID="adige20040907_id405381-E1" TYPE="PER" SUBTYPE="Group" CLASS="SPC">
    <entity_mention ID="adige20040907_id405381-E1-1" TYPE="NOM" PRIMARY="true" ME
      <extent>
        <charseq START="235" END="245">dei parenti</charseq>
      </extent>
      <head>
        <charseq START="239" END="245">parenti</charseq>
      </head>
    </entity_mention>
  </entity>
</document>
```



```

accomagnato __nil__ 39 198 210 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
le __nil__ 40 211 213 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
lacrime __nil__ 41 214 221 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
e __nil__ 42 222 223 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
lo __nil__ 43 224 226 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
strazio __nil__ 44 227 234 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
dei __nil__ 45 235 238 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 B-EXT+adige20040907_id405381-E1#PER#Group#SPC#
parenti __nil__ 46 239 246 - adige20040907_id405381.txt 20040907 Attualita B-HEA+adige20040907_id405381-E1#PER#Group#SPC#a
- __nil__ 47 247 248 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
hanno __nil__ 48 249 254 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
ricevuto __nil__ 49 255 263 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
l' __nil__ 50 264 266 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
ultimo __nil__ 51 266 272 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0
saluto __nil__ 52 273 279 - adige20040907_id405381.txt 20040907 Attualita 0 0 0 0

```

Development set, using CoNLL scorer:

Measure		1st level classifier	2nd level classifier
General	Precision	85.93%	50.87%
	Recall	81.61%	36.68%
	FB1	83.71%	42.63%
NAM	Precision	87.29%	44.09%
	Recall	88.71%	33.73%
	FB1	87.99%	38.23%
NOM	Precision	86.75%	56.60%
	Recall	75.42%	39.30%
	FB1	80.69%	46.39%
PRO	Precision	79.81%	0.00%
	Recall	72.44%	0.00%
	FB1	75.95%	0.00%

Measure		1st level classifier	2nd level classifier
General	Precision	81.12%	50.70%
	Recall	77.07%	36.43%
	FB1	79.04%	42.40%
GPE	Precision	83.28%	46.15%
	Recall	82.48%	31.58%
	FB1	82.88%	37.50%
LOC	Precision	75.73%	0.00%
	Recall	58.65%	0.00%
	FB1	66.10%	0.00%
ORG	Precision	75.98%	36.96%
	Recall	69.91%	19.54%
	FB1	72.82%	25.56%
PER	Precision	83.26%	54.22%
	Recall	80.29%	43.11%
	FB1	81.75%	48.03%

<b>EMD Task</b>	<b>Measure</b>	<b>Result</b>
	Value	65.7%
	Precision	78.1%
	Recall	74.1%
	F-measure	76.1%

<b>Local Entity Detection and Recognition Task</b>	<b>Measure</b>	<b>Result</b>
	Value	36.7%
	Precision	78.5%
	Recall	61.1%
	F-measure	68.7%

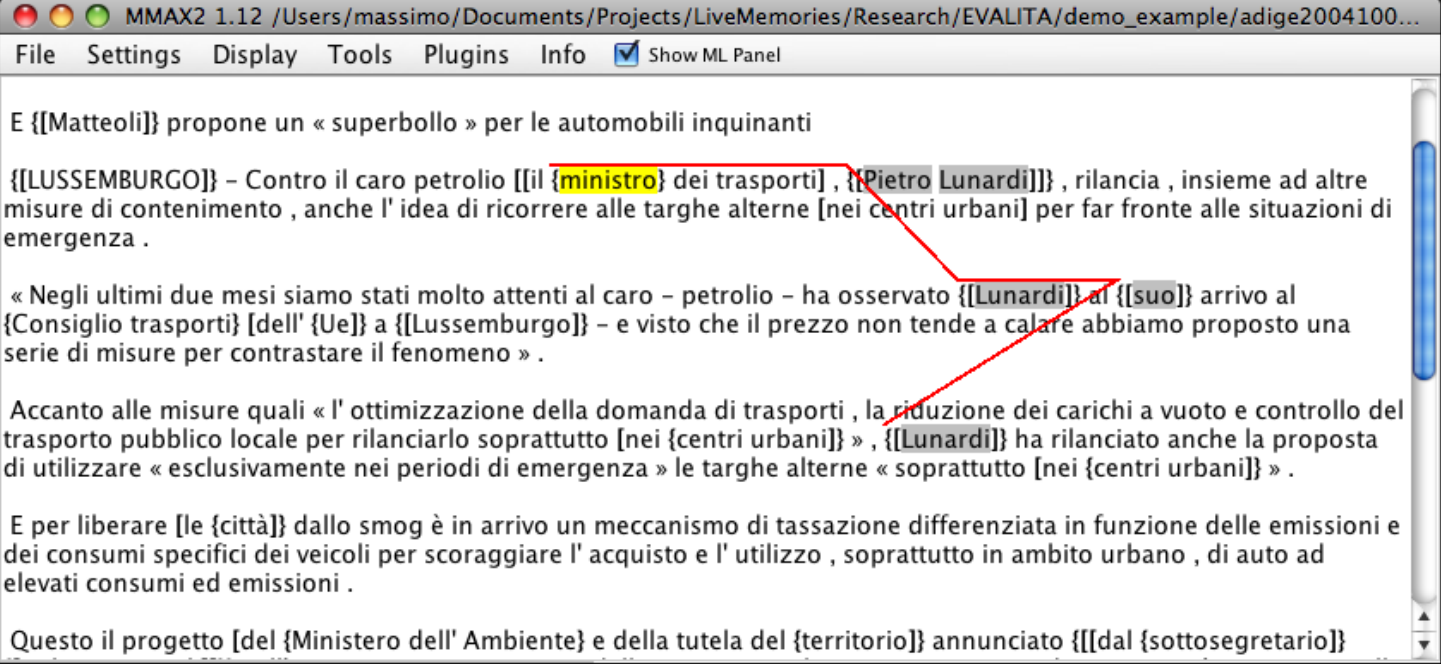
- An anaphora resolution toolkit developed at the 2007 Johns Hopkins ELERFED workshop
- Designed to allow experimentation with anaphora resolution:
  - Modular architecture
  - Variety of preprocessing
  - Declarative specification of configuration
- Java-based, open source ([www.bart-coref.org](http://www.bart-coref.org))

# ITALIAN ANAPHORA RESOLUTION IN BART

- Developed a LANGUAGE PLUGIN isolating language-specific aspects
  - E.g., aliasing: S.p.A., D.ssa
- Modified preprocessing to work with either mention extractor (DeMention) or with parsers (Italian version of MALT parser trained on TUT)
- Configuration: Soon et al 2001



# LINKING ENTITY MENTIONS WITH ITAL-BART



E {{Matteoli}} propone un « superbollo » per le automobili inquinanti

{{LUSSEMBURGO}} - Contro il caro petrolio [[il {ministro} dei trasporti], {{Pietro Lunardi}}], rilancia , insieme ad altre misure di contenimento , anche l' idea di ricorrere alle targhe alterne [nei centri urbani] per far fronte alle situazioni di emergenza .

« Negli ultimi due mesi siamo stati molto attenti al caro - petrolio - ha osservato {{Lunardi}} al {{suo}} arrivo al {Consiglio trasporti} [dell' {Ue}] a {{Lussemburgo}} - e visto che il prezzo non tende a calare abbiamo proposto una serie di misure per contrastare il fenomeno » .

Accanto alle misure quali « l' ottimizzazione della domanda di trasporti , la riduzione dei carichi a vuoto e controllo del trasporto pubblico locale per rilanciarlo soprattutto [nei {centri urbani}] » , {{Lunardi}} ha rilanciato anche la proposta di utilizzare « esclusivamente nei periodi di emergenza » le targhe alterne « soprattutto [nei {centri urbani}] » .

E per liberare [le {città}] dallo smog è in arrivo un meccanismo di tassazione differenziata in funzione delle emissioni e dei consumi specifici dei veicoli per scoraggiare l' acquisto e l' utilizzo , soprattutto in ambito urbano , di auto ad elevati consumi ed emissioni .

Questo il progetto [del {Ministero dell' Ambiente} e della tutela del {territorio}] annunciato {{{dal {sottosegretario}}}

The screenshot shows a text editor window with a menu bar (File, Settings, Display, Tools, Plugins, Info) and a 'Show ML Panel' checkbox. The text contains several entity mentions in curly braces, some of which are linked by a red line. The word 'ministro' is highlighted in yellow. The text discusses transportation measures and mentions 'Lussemburgo', 'Pietro Lunardi', and 'Lunardi'.

Thank you