My Geography Manual is Now Linked with the Web and the Real World

Dan Cristea

“Alexandru Ioan Cuza” University of Iași, Faculty of Computer Science
Romanian Academy, Institute for Computer Science
dcristea@info.uaic.ro
Acknowledgements...

MappingBooks – Enter the book!
Evade from the book in the virtual and real world!

a PN-II project (July 2014 – September 2017) financed by the Romanian Ministry of Education and Research
I like traveling and reading...
Going out of the book...
I need help to remember all kinship relations between characters.
Characters in *Forsyte Saga*

- **The old Forsytes**
  - Ann, the eldest of the family
  - Old Jolyon, the patriarch of the family, having made a fortune in tea
  - James, a solicitor, married to Emily, a most tranquil woman
  - Swithin, James's twin brother with aristocratic pretensions; a bachelor
  - Roger, "the original Forsyte"
  - Julia (Juley), a fluttery dowager; Mrs. Septimus Small
  - Hester, an old maid
  - Nicholas, the wealthiest in the family
  - Timothy, the most cautious man in England
  - Susan, the married sister

- **The young Forsytes**
  - Young Jolyon, Old Jolyon's artistic and free-thinking son, married three times
  - Soames, James and Emily's son, an intense, unimaginative and possessive solicitor, married to the unhappy Irene, who later marries Young Jolyon
  - Winifred, Soames's sister, one of the three daughters of James and Emily, married to the foppish and lethargic Montague Dartie
  - George, Roger's son, a dyed-in-the-wool mocker
  - Francie, George's sister and Roger's daughter, emancipated from God

- **Their children**
  - June, Young Jolyon's defiant daughter from his first marriage; engaged to an architect, Philip Bosinney, who becomes Irene's lover
  - Jolly, Young Jolyon's son from his second marriage; dies of enteric fever during the Boer Wars
  - Holly, Young Jolyon's daughter from his second marriage, to June's governess
  - Jon, Young Jolyon's son from his third marriage, to Irene, Soames's first wife
  - Fleur, Soames's daughter from his second marriage, to a French Soho shopgirl Annette; Jon's lover; later marries a baronet, Michael Mont
  - Val, Winifred and Montague's son; fights in the Boer Wars; marries his cousin Holly
  - Imogen, Winifred and Montague's daughter

- **Others**
  - Parfit, Old Jolyon's butler
  - Smithier, Aunts Ann, Juley and Hester's housekeeper
  - Warmson, James and Emily's butler
  - Bilson, Soames's housemaid
  - Prosper Profond, Winifred's admirer and Annette's lover
Bring back the book in the hands of children!

• What do youngsters keep in their hands in our times?
  – Tablets
  – Kendamas
  – Books?...

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Linguistics Linked Open Data (LLOD)

a subfield of Natural Language Processing

- Develop techniques able to \textbf{decipher the semantic content of texts}
  - narrative lines (e.g. what happens and when)
  - semantic relations between entities (e.g. genealogical trees, spatial and temporal relations)
  - statistics about entities (# mentions $\rightarrow$ salience, etc.)
  - summaries (general, focused on characters)
Linguistics Linked Open Data (LLOD)

- Generation of ontologies from collections of scientific works
  - applications that “read” science books and formalize concepts and their instances

- Intelligent documentary search
  - Personalized assistants of a research activity
Entity linking

• Challenges in entity linking:
  – name variations
  – ambiguities
  – first mentions
  – reference chains
Linking entities internally
The ‘QuoVadis’ corpus

Chapter I

PETRONIUS woke only about midday, and as usual greatly wearied. The evening before he had been at one of Nero's feasts, which was prolonged till late at night. For some time his health had been failing. He said himself that he woke up benumbed, as it were, and without power of collecting his thoughts. But the morning bath and careful kneading of the...
A corpus semantic entities and relations

• Type of entities:
  – persons
  – gods
  – groups of persons and gods
  – body parts

• Semantic relations among entities of these types
Relations

- Anaphoric relations: co-referential;
- Non-anaphoric relations:
  - kinship;
  - affective;
  - social.
Anaphoric relations

- coref
- coref-interpret
- member-of, has-as-member (inverse)
- isa, class-of (inverse)
- part-of, has-as-part (inverse)
- subgroup-of, has-as-subgroup (inverse)
- has-name, name-of (inverse)

Example:

Lygia\textsubscript{1} was unable to answer, for weeping seized her\textsubscript{2} anew. Acte gathered the maiden\textsubscript{3} to her bosom, and strove to calm her\textsubscript{4} excitement.

Kinship relations

- parent-of
- child-of (inverse of parent-of)
- grandparent-of and grandchild-of (inverse)
- sibling (symmetrical)
- ant-uncle-of, nephew-of (inverse relation)
- cousin-of (symmetrical)
- spouse-of (symmetrical)
- unknown

Example:
"Pardon me, Lygia. For me thou art [<the daughter> [of a king]₂]₁ and [<the adopted child> [of Plautius]₄]₃."

Social relations

• superior-of
• inferior-of
• in cooperation-with
• colleague-of
• in competition-with
• opposite-to

Example:

[Petronius]\_1…but to [his]\_2 misfortune [he]\_3 <surpassed in conversation> [Caesar himself]\_4, hence [he]\_5 roused [his]\_6 jealousy.

[3] in competition-with [4];
Affective relations

- love
- loved-by
- hate
- hated by
- upset
- friendship
- worship

Example:
Vinicius entered Lygia's dungeon and remained there till daylight...Both changed by degrees into sad souls <in love> with \([each]_1 \ [other]_2\).

Relations

- Anaphoric: coref

\[ \begin{align*}
\text{John} & \quad \text{met Maria on the ski slope.} \\
\text{He} & \quad \text{raced her.}
\end{align*} \]
Relations

• Anaphoric: coref

John met Maria on the ski slope. He raced her.
Arguments and triggers in relations

• Kinship: parent-of
Arguments and triggers in relations

• Social: inferior-of

Cesar's principal courtiers

destination

trigger

source
Arguments and triggers in relations

• Affective: worship

Lygia dropped on her knees to implore someone else.
EnAAes

Petronius... 

Vinicius was

the son of his oldest sister, who years before had married his father, a man of consular dignity from the time of Tiberius.
Anaphoric relations: coref

Petronius,

Vinicius was the son of his oldest sister,

who years before had married his father,

a man of consular dignity from the time of Tiberius.
Petroniu...  

Vinicius was the son of his oldest sister, who years before had married his father, a man of consular dignity from the time of Tiberius.
Anaphoric relations: coref

Petronius...

Vinicius was

the son of his oldest sister,

who years before had married his father,

a man of consular dignity from the time of Tiberius.

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Anaphoric relations: class-of

Petronius...

Vinicius was

the son of his oldest sister,

who years before had married his father,

a man of consular dignity from the time of Tiberius.
Kinship relations: sibling

Petronius...

Vinicius was the son of his oldest sister, who years before had married his father, a man of consular dignity from the time of Tiberius.
Petronius...

Vinicius was the son of his oldest sister, who years before had married his father, a man of consular dignity from the time of Tiberius.
Kinship relations: parent-of

Petronius...

Vinicius was

the son of his oldest sister,

who years before had married his father,

a man of consular dignity from the time of Tiberius.
Kinship relations: spouse-of

Petronius...

Vinicius was the son of his oldest sister, who years before had married his father, a man of consular dignity from the time of Tiberius.
Social relations: inferior-of

Petronius...

Vinicius was

the son of his oldest sister,

who years before had married his father,

a man of consular dignity from the time of Tiberius.
General statistics over the corpus

• 7,281 sentences
• 146,822 tokens, punctuation included
• 171,029 tokens summed up under all relations
• 24,636 entity mentions
• 22,301 referential relations
• 755 AKS relations (Affective + Kinship + Social)
• 752 triggers
Example: affective relations love and worship
Example: affective relations *fear-of* and *hate*
Vinicius’ links with other characters
Semantic relations involving Vinicius

Vinicius' Relationships

- superior-of, inferior-of (35%)
- love, rec-love (23%)
- friend-of (8%)
- hate (5%)
- worship (3%)
- nephew-of (4%)
- child-of (10%)
- aunt/uncle-of (1%)
- parent-of (1%)
- colleague-of
- sibling (2%)
- spouse-of (6%)

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Marcus Vinicius fiu fiul surorii sale mai mari,


Tiberiu

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,

Marcus Vinicius, sănătos, fiul său mai mari, fiul său mai mari,
Linking entities externally
MappingBooks – a bird’s view
MappingBooks

• A *MappedBook* is a book connected with locations/events in the virtual and real world and sensitive to the instantaneous location of the reader (as seized by the telephone/tablet).

• The information made available could possibly be different depending on the moment and the place of the reader.
MappingBooks

- Multi-dimensional mash-ups combining textual and geographical data
- Spot book mentions of entities (persons and locations) and link them in the virtual world
- Make heavy use of entity linking techniques
- Easy to handle interface for young readers
The application

1) Connects mentions of entities (nominal groups) => one entity = a chain of coreferential mentions
2) The knowledge base does not include any apriory records about entities => starts from scratch
3) Identifies geographical relations (distances, positions, proximities, intersections, etc.)
4) Texts, for the time being: geography manuals
Entity types in MB

- Type PERSON
- Type LOCATION
- Type ORGANISATION
- Type URL
- Type TIMEX
Textual realisation of entities

• Syntactic realisation: NPs (proper nouns, common nouns, adjectives, complement PPs; but NO relative clauses)

• Characterised by distinctive heads
  – \([\text{the house on the } [\text{mountain}]]\)

• If intersected \(\rightarrow\) imbricated
  – \([\text{the museum } [\text{Grigore Antipa}]]\)
Processing features

• The capacity to see a text different than a string of letters
  – sentence splitting
  – tokenisation
  – POS-tagging
  – lemmatisation
  – NP chunking
  – anaphora resolution
Processing features

• Know who’s who
  – recognise names and types
  – disambiguate names
  – recognise an entity in the text even if mentioned by a common noun or a pronoun
  – use an ontology of types

NAME ENTITY RECOGNITION

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Processing features

• What virtual world entities are mentioned in the book?
  – link textual mentions of entities in the virtual world
  – decide what virtual info would be relevant to user
  – employ multiple sources

ENTITY CROWLING
Processing features

• Fetch, process and make use of geo-data
  – Geographic Information Systems (GIS)
  – geographic layers
Processing features

• Trace on a map a spatial relation described in the book
  – spatial relations detection in text
  – use Google Maps-like geo-strata (actually we procured our own maps)
  – trace locations and paths on maps

RELATIONS DETECTION
MAPS&TRAJECTORIES

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Processing features

• Know where I am
• What real world entities are in my proximity
  – detection of my position
  – computation of distances from the mentioned places
  – signalling “interesting” locations in proximity

DElCE INFO
Processing features

• Mix images with generated info
  – locate the position of the user (GPS)
  – Sense the orientation of the camera (compass)
  – process images => segment, contours, recognition
  – decide info to be displayed
Processing features

- Attractive user interfaces
  - analyse use cases
  - design dedicated user interfaces
  - accommodate on the screen a segment of text, a map, user’s position, web info, etc.

INTERFACES

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
Processing features

• Client-server
  – user’s Portrait
  – the databases
  – standards and communication protocols
Other issues...

• RESOURCES
  – find the texts
  – clear IPR
  – perform annotation
  – find other relevant linguistic data
TA = Text Analytics
NER = Name Entity Recognition
EC = Entity Crowling
RD = Relations Detection
GEO = Geography
M&T = Maps and Trajectories

AR = Augmented Reality
DEV = Device Info
INT = Interfaces
RES = Resources
M&E = Management and Evaluation

IDAACS-2017, University Politehnica of Bucharest, 22 Sept.
What else could be added?

Networking Readers

• Using semantic and geographical links to form social communities of readers
  – if books “subscribed for” declared visible =>
    • co-readers of B (book)
  – if “instantaneous location” also declared visible =>
    • co-readers of B AND actual co-proximity of L (location)
    • co-readers of B AND co-track of T (trajectory)
Networking Readers: enhance e-Books reading experience

• Easy to imagine other ways to form communities rooted in readings
  – intersect **common readings** and **attended places** with **levels of friendship** reported by other social media, like Facebook or Twitter
  – **real-world events** and **entities mentioned** in a book associated with **real-world locations** and particular **moments of the year/day**
  – portraying the user (from accessible social media and habits of MB behavior) and matching
Usage examples

- I visit a city with the traveling guide in my hand...
  - places of interest, routes, are reordered depending on my instantaneous position
Usage examples

- I am a school boy, in the train going from Braşov to Sibiu...
  - if I open my tablet and head it towards the left side window of the train, I will see arrows showing the picks of the Făgăraş mountains, exactly as in the Geography manual
Augmented reality
Usage examples

- I am in Paris for the 3\textsuperscript{rd} time...
  - but only now my MB Lonely Planet guide signals me this temporary exhibition opened in the Pyramid
Towards... live books

• Multidimensional artefacts that combine textual, geographical, temporal, etc. data
• Evidence mentions of persons, locations...
• Links sensible at:
  – the context of the mention in book
  – the location of the reader
  – the moment of the lecture
  – the personality and preferences of the reader
Real or virtual?
Troubling questions...

• To what degree should the technology interfere with the act of literary creation?

• Virtual reality techniques are at our touch
  – still, which of the description of perception mentioned in a book should be added to the text?

• We do not want to rebuild in virtual sensations, images, sounds...
  – we only want to help the reader with complementary information
MappingBooks: more acknowledgements

• Undergrad. students from the Faculty of Computer Science, who in the first term of the univ. year 2013-2014 have build the MB prototype as a term project in the AI course.

• My colleagues: Ionuț Pistol, Daniela Gifu (Fac. CS), Mihai Niculiță (Fac. Geography) in UAIC.

• Partners in the MB project: Univ. “Ștefan cel Mare” Suceava, SIVECO – Bucharest
References


Thank you!