Using NLP and Ontologies for Notary Document Management Systems

Flora Amato, Antonino Mazzeo,
Antonio Penta and Antonio Picariello

Dipartimento di Informatica e Sistemistica
Universitá di Napoli “Federico II”, Italy
{flora.amato, mazzeo, a.penta, picus}@unina.it

September 01, 2008

Outline

1. Dematerialization
   - The Dematerialization Activity
   - The legal information system
   - The System Functionalities
   - The System Architecture
2. Knowledge Modeling
   - Structural Ontology
   - Domain Ontology
3. Procedure
   - Document Partition
   - Ontology Fragment $\rightleftharpoons$ Document Segment
   - Ontology Population

Outline

1. Dematerialization
   - The Dematerialization Activity
   - The legal information system
   - The System Functionalities
   - The System Architecture
2. Knowledge Modeling
   - Structural Ontology
   - Domain Ontology
3. Procedure
   - Document Partition
   - Ontology Fragment $\rightleftharpoons$ Document Segment
   - Ontology Population

Outline

1. Dematerialization
   - The Dematerialization Activity
   - The legal information system
   - The System Functionalities
   - The System Architecture
2. Knowledge Modeling
   - Structural Ontology
   - Domain Ontology
3. Procedure
   - Document Partition
   - Ontology Fragment $\rightleftharpoons$ Document Segment
   - Ontology Population

The Dematerialization Activity

1. In almost all legal traditional activities, most of the processes are characterized by the presence of hard paper documents that have to be properly managed: processed, archived and prepared for long term preservation
   - An intense and extensive dematerialization activity is still necessary
2. **The dematerialization process implies the application of syntactic-semantic methodologies in order to automatically transform the unstructured or sometimes semi-structured document into a formally structured, machine readable records.**

The legal information system

Proposal

The core of our legal information system:

- manages a variety of hard paper documents,
- automatically transforms them into *RDF* statements for suitable:
  - indexing,
  - retrieval,
  - long term preservation.

Although we describe a general architecture that can be used for several legal application domains, our system is particularly suitable for the notary realm.

A Running Example: Notary Documents

- Let us suppose to analyze a *buying act* some document contents are:
  - the description of buying or selling a property, such as houses, pieces of lands,
  - the identify of people involved,
  - the official code and data

REPERTORIO 55623 RACCOLTA 2445327
COMPRAVENDITA
L'anno duemilasei il giorno 28 marzo in Treviso, nella Sede municipale di Ca' Sugana
in via Municipio n. 16, davanti a me dottor Carmine Preciso,
notaio iscritto al Collegio del Distretto Notarile di Trieste omessa
l'assistenza di testimoni per avervi gli infrascritti comparenti, aventi i
requisiti di legge, d'accordo tra di loro e con il mio consenso, espressamente
rinunziato,
vi sono da una parte:
1) Andrea Arno, nato a Napoli il 17/8/58, domiciliato in via Campi Elisei n.
23 in Treviso
dall'altra parte:
2) Luciana Lana, nata a Pozzuoli (Na) il 12/7/78 e residente a Torino in
via Guglielmo I numero 52,
Ciò premesso,
confermato e ritenuto parte integrante e contestuale dell'atto, i comparenti
convengono e dichiarano quanto segue:
Art. 1) Il signor Arno Andrea proprietario, come sopra rappresentato,
vende e trasferisce a favore della signorina Luciana Lana, che
accetta ed acquista, l'immobile appresso indicato:
Catasto Terreni - Comune di Treviso - Area di enti urbani promiscui
Fg. 27 - particella 722 - ente urbano - Ha 0.04.40
Pagina 4 di 9

The Running Example: The implicit structure of Notary Document

Such document is generally composed by:

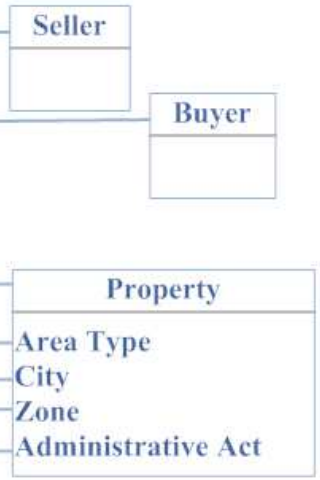
- an *introduction part* containing the captions,
- a *biographical part* containing of the individuals involved in the buying act,
- a *property part* that is a section containing data about the property and a sequence containing several rules regulating the sales contract.

The main aims are :

- to design models and algorithms able :
  - to detects the several sections containing relevant information in order
  - to transform the unstructured information within the retrieved section into a structured document.
- To apply that analysis on a real system.

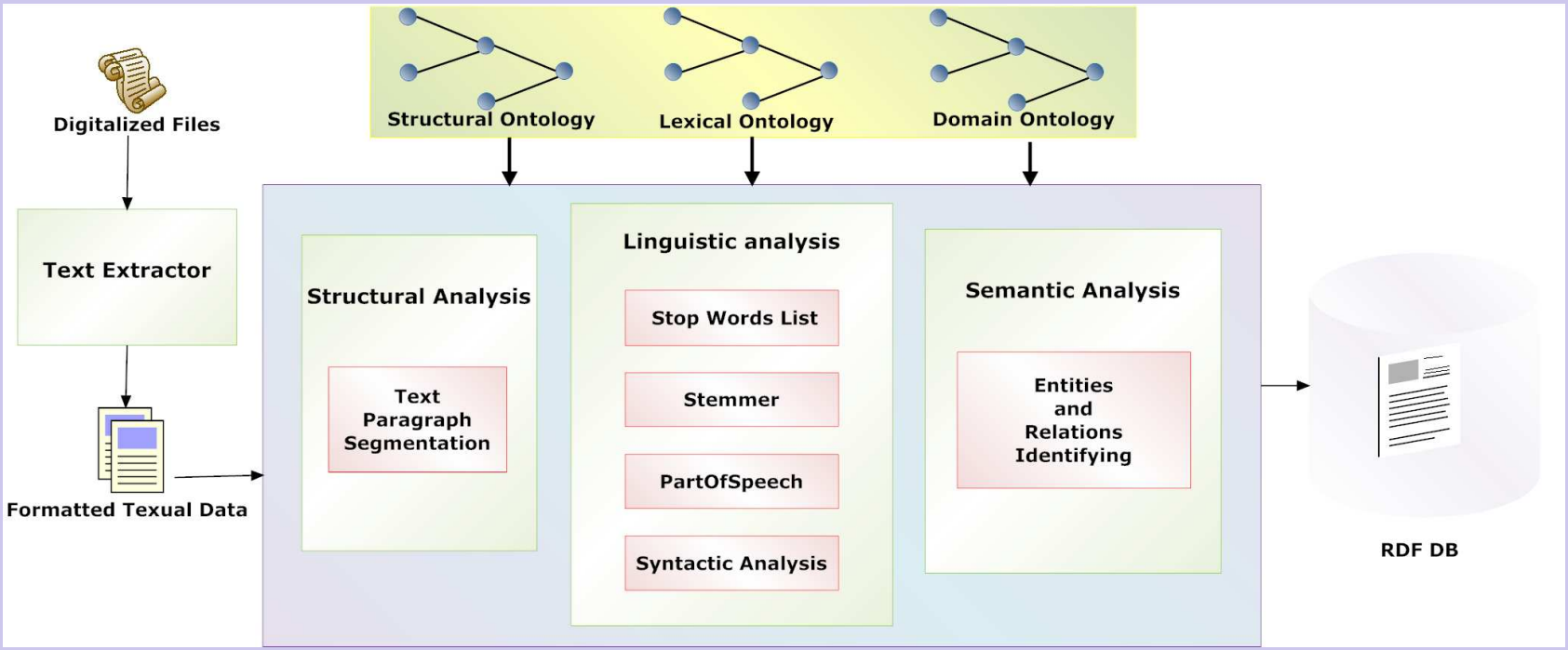
The Running Example:An example of Document Processing

REPERTORIO 55623 RACCOLTA 2445327
COMPRAVENDITA
L'anno duemilasei il giorno 28 marzo in Treviso, nella Sede municipale di Ca' Sugana
in via Municipio n. 16, davanti a me dottor Carmine Preciso,
notaio iscritto al Collegio del Distretto Notarile di Trieste omessa l'assistenza di testimoni per avervi gli infrascritti comparenti, aventi i requisiti di legge, d'accordo tra di loro e con il mio consenso, espressamente rinunziato,
vi sono da una parte:
1) Andrea Arno, nato a Napoli il 17/8/58, domiciliato in via Campi Elisei n. 23 in Treviso
dall'altra parte:
2) Luciana Lana, nata a Pozzuoli (Na) il 12/7/78 e residente a Torino in via Guglielmo I numero 52,
Ciò premesso,
confermato e ritenuto parte integrante e contestuale dell'atto, i comparenti convengono e dichiarano quanto segue:
Art. 1) Il signor Arno Andrea proprietario, come sopra rappresentato, vende e trasferisce a favore della signorina Luciana Lana, che accetta ed acquista, l'immobile appresso indicato:
Catasto Terreni - Comune di Treviso - Area di enti urbani promiscui
Fg. 27 - particella 722 - ente urbano - Ha 0.04.40
Pagina 4 di 9
esattamente individuato negli elaborati planimetrici che in fotocopia previo esame e sottoscrizione delle parti e di me notaio si allegano all'atto.
Art. 2) L'immobile confina a Nord - Ovest con Via Barberia, a Sud con il mapp. 726 e a Est con i 723-725.
Art. 3) L'immobile è pervenuto al patrimonio comunale a seguito donazione del Vice Re d'Italia con rescritto del 4 luglio 1812. La donazione è stata partecipata al Podestà del Comune di Treviso dal Prefetto del Tagliamento con lettera del 28/7/1812 n. 16572.
Art. 4) Quanto sopra viene compravenduto, considerato a corpo, nello stato e grado in cui attualmente si trova, ed in quello di diritto in cui è pervenuto ed è stato fino ad oggi posseduto dalla parte venditrice, in virtù del titolo e del possesso, con accessioni e pertinenze, oneri e diritti inerenti, azioni, ragioni e servitù e passive, apparenti e non apparenti, ove esistenti, se ed in quanto legalmente costituite, ivi compresa la servitù di veduta a favore del



The system architecture

System Overview



Knowledge Modeling:Structural Ontology

- In the legal domain, almost all the documents are still written using natural languages.
- Nevertheless, the unstructured form of document follows a well determined sequence:
  - in a notary act, for example, the notaries use a certain subset of natural language and in addition they use a certain pre-defined structure, that is codified by laws or normative rules.
- For these reason, we say that our legal realm manages *semi-structured documents* written in a simplified natural language.

Structural Ontology

On this basis we codify the information on the structure, extracted by domain expert, in the Structural Ontology

Knowledge Modeling: Domain Ontology

Domain Ontology

The Domain Ontology contains the concepts and the relations to be extracted

- This Ontology is designed by the help of expert domain.
- **The whole ontology is divided in fragment**

Ontology Fragment

Every fragment contains informations about some

- concepts of interest and
- his proprieties

Act Segmentation

Using information codified in the Structural Ontology, we subdivide the whole document in segments,
The text of the act is portioned in different ways, using several partition rules that are dependent from:

1. normative prescriptions;
2. tradition of single notary schools;
3. common use of the single notary.

Segments in which the whole act can be segmented

REPERTORIO 55623 RACCOLTA 2445327
COMPRAVENDITA
L'anno duemilasei il giorno 28 marzo in Treviso, nella Sede municipale di Ca' Sugana
in via Municipio n. 16, davanti a me dottor Carmine Preciso,
notaio iscritto al Collegio del Distretto Notarile di Trieste omessa
l'assistenza di testimoni per avervi gli infrascritti comparenti, aventi i
requisiti di legge, d'accordo tra di loro e con il mio consenso, espressamente
rinunziato,
vi sono da una parte:
1) Andrea Arno, nato a Napoli il 17/8/58, domiciliato in via Campi Elisei n. 23 in Treviso
dall'altra parte:
2) Luciana Lana, nata a Pozzuoli (Na) il 12/7/78 e residente a Torino in
via Guglielmo I numero 52,
Ciò premesso,
confermato e ritenuto parte integrante e contestuale dell'atto, i comparenti
convengono e dichiarano quanto segue:
Art. 1) Il signor Arno Andrea proprietario, come sopra rappresentato,
vende e trasferisce a favore della signorina Luciana Lana, che
accetta ed acquista, l'immobile appresso indicato:
Catasto Terreni - Comune di Treviso - Area di enti urbani promiscui
Fg. 27 - particella 722 - ente urbano - Ha 0.04.40
Pagina 4 di 9
esattamente individuato negli elaborati planimetrici che in fotocopia
previo esame e sottoscrizione delle parti e di me notaio si
allegano all'atto.
Art. 2) L'immobile confina a Nord - Ovest con Via Barberia, a Sud
con il mapp. 726 e a Est con i 723-725.
Art. 3) L'immobile è pervenuto al patrimonio comunale a seguito
donazione del Vice Re d'Italia con rescritto del 4 luglio 1812. La
donazione è stata partecipata al Podestà del Comune di Treviso dal
Prefetto del Tagliamento con lettera del 28/7/1812 n. 16572.
Art. 4) Quanto sopra viene compravenduto, considerato a corpo,
nello stato e grado in cui attualmente si trova, ed in quello di diritto in
cui è pervenuto ed è stato fino ad oggi posseduto dalla parte
venditrice, in virtù del titolo e del possesso, con accessioni e
pertinenze, oneri e diritti inerenti, azioni, ragioni e servitù e
passive, apparenti e non apparenti, ove esistenti, se ed in quanto
legalmente costituite, ivi compresa la servitù di veduta a favore del
Comune di Treviso nel sottoportico della corte di Ca' dei Ricchi,
costituita con atto rep. 43232 del 13/09/1995 a rogito del Segretario
Pagina 5 di 9
Generale del Comune di Treviso dott. Giuseppe Sorge.

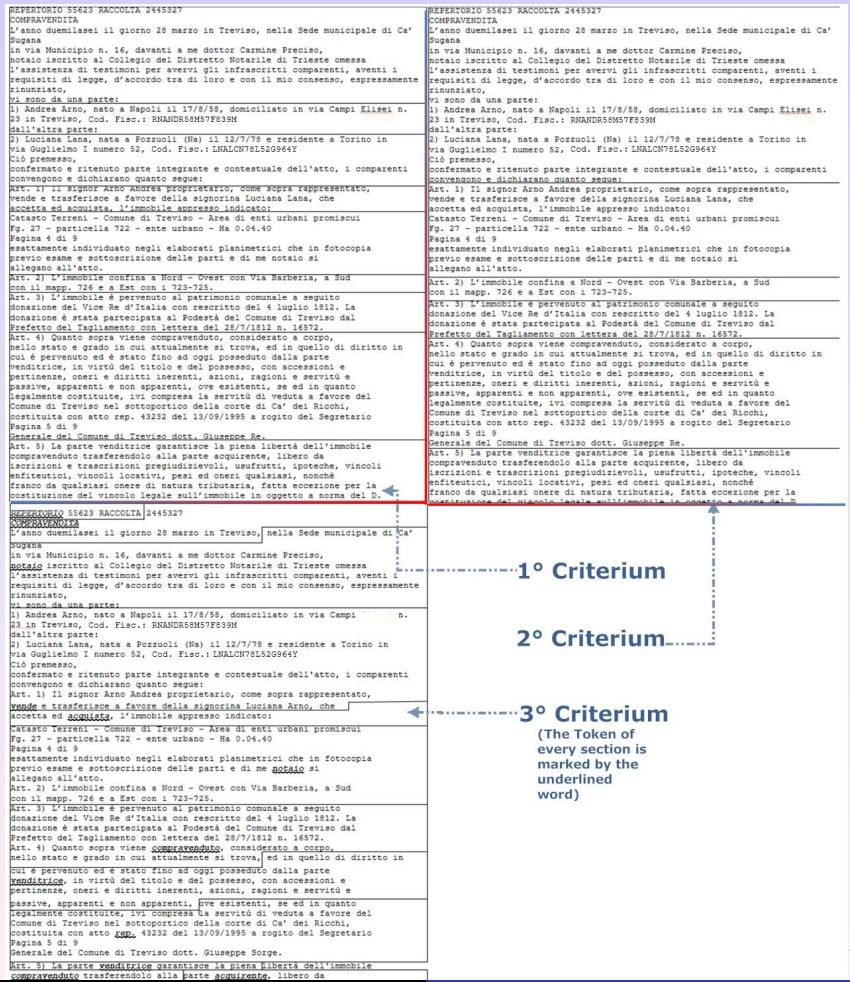
Introduction Part

Biographical Data Part

Property Data Part

Rules Part

Example of application of Different Partition Criteria on the same Act



Association between Ontology Fragment and Document Segment

- We want to fill the ontology fragments with the instances extracted from the document
- In order to associate the most suitable act part with ontology fragments, we apply a scoring criterion realized comparing the **pattern** extracted from the text segment with the concept contained in the ontology fragment.
- In order to identify the concept in the text, we use synonyms sets obtained by querying ItalWordnet on the significant word in the text
  - significant words are identified in the Linguistic Analysis (Stop Word List and Part of Speech)

Ontology Population

- In order to populate the ontology fragment we apply the rules codified in the domain ontology
- Such rules contain and use
  - the expert domain knowledge and
  - NER functionality to recognize the entity

Rule Example

We report a simple rule, that is triggered when a sentence contains a synonymous of the the verb “sell” or one of its conjugation.
To identify a seller in the document we apply the rule, in which x is the seller

$$\texttt{Seller(x)} \Leftarrow \texttt{Person(x)} \wedge \texttt{Sell(x,y)} \wedge \texttt{Object(y)}$$

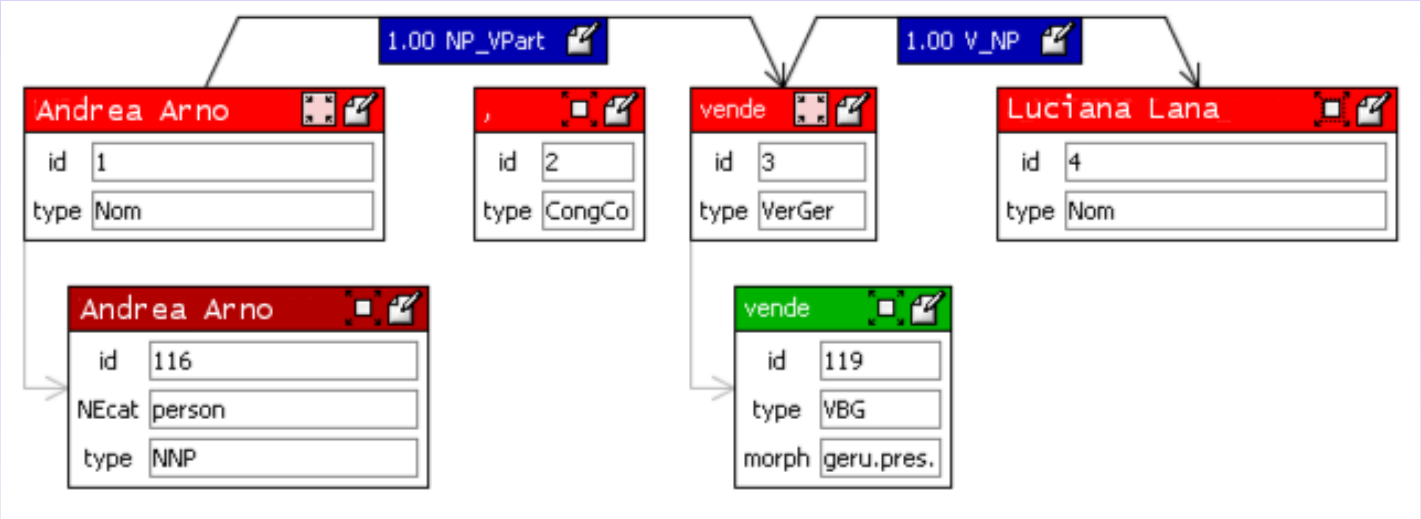
Example: Rule Application

In the phrase

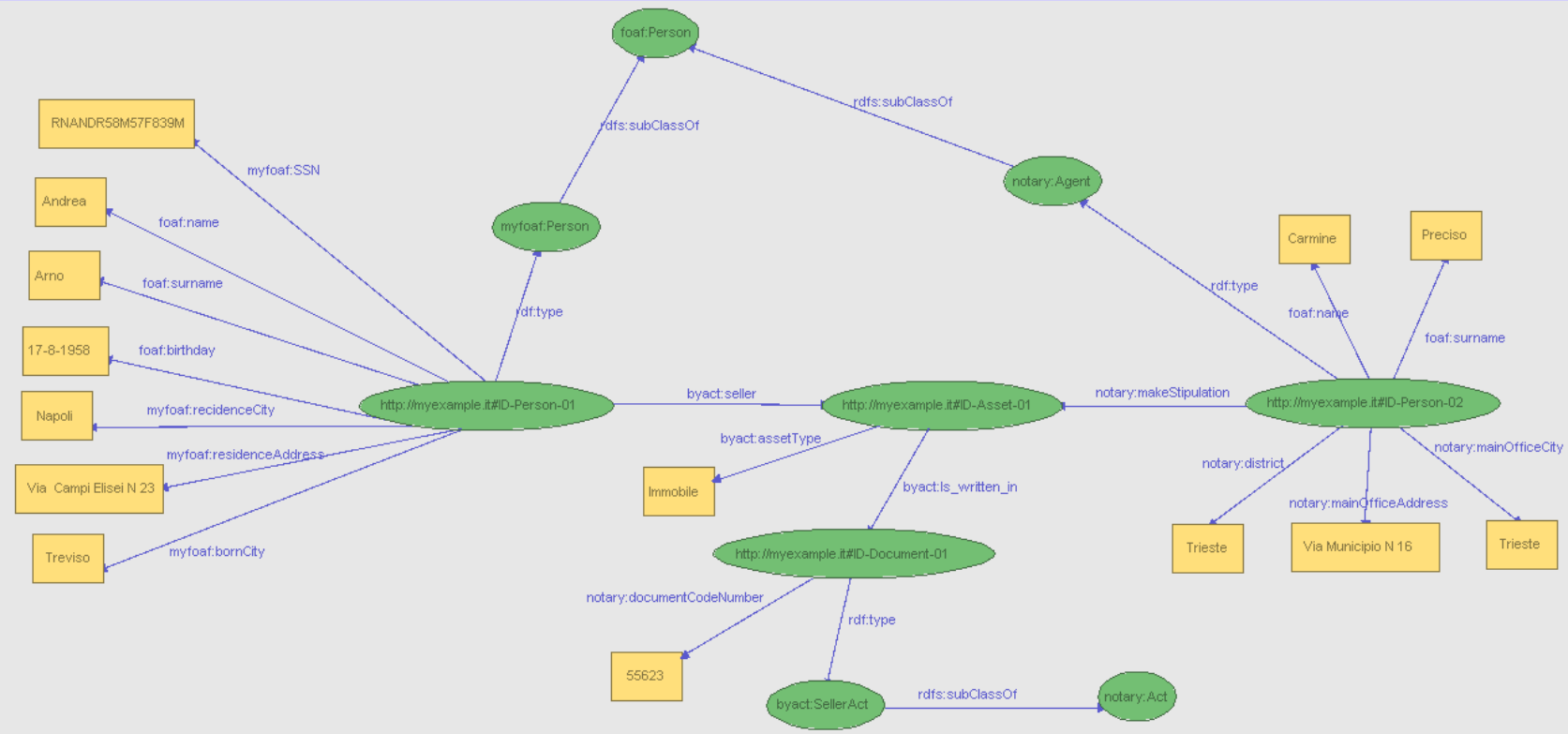
Andrea Arno sells the house to Luciana Lanna

- The NER module identifies that “Andrea Arno” and “Luciana Lanna” are persons
- The Lexical Ontology identifies the set of synonyms for the word “house” (identifying it is an object)
- The Lexical Ontology identifies the set of synonyms for the verb “Sell”
- The presence of the sell concept triggers the rule, codified in the domain ontology:
- $\texttt{Seller(x)} \Leftarrow \texttt{Person(x)} \wedge \texttt{Sell(x,y)} \wedge \texttt{Object(y)}$

Example:Syntactic Rules



A section of RDF graph extracted from a notary act



Conclusion and Future Works

- Conclusion:
  - We have presented a general system as the core of an e-gov information system in the notary domain.
  - We propose to use different levels of ontology and of NLP techniques in order : to transform semi-structured documents into structured *RDF* triples
- Future Works:
  - Improving of the domain and structure ontology
  - Using of Semantic Wiki in collaboration with National Notary Council for the assessment of the concepts and relations inside the structural and domain ontology
  - We are planning a large experimentations in the Italian notary domain.