Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Exploration of documents concerning Foundlings in Fafe along XIX Century

João Costeira Faria Gomes

Universidade do Minho

2022

Outline

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

1 Contextualization
2 Motivation
3 Objectives
4 Documents
5 Architecture
6 Implementation
- Models and Persistence
- Entity Extraction
- Relation Extraction
- Pipeline Integration

7 Example
8 Conclusion

Contextualization

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- The Foundling Wheel was introduced to combat the anonymous child abandonment problematic.

- Originally spread across Europe during the XIII century.

- Measure implemented to reduce the high mortality rate, consequent to the anonymous abandonment.

Contextualization

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- Wheel and Bell components provide a safe abandonment, without compromising the anonymity of the perpetrators.

- These public institutions raised the foundling population, additional support provided by hiring nurses and wet nurses.

- The Foundling Wheel is defunct, replaced by Hospices and family based support.

Contextualization

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- The Municipal Archive of Fafe preserves documents from these historical institutions.

Case Study

The case study consists in the analysis of archived Foundling Wheel and Hospices documents, dated to the nineteenth century, from the Northern region of Portugal.

Motivation

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- Preserve and disseminate information contained in these historical documents.

- Historical contextualization to comprehend the effectiveness of these programs and the life conditions provided.

- Current debate towards the reintroduction of anonymous child abandonment mechanisms.

Objectives

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Research Hypothesis

From the identification of concepts and respective relations over the Foundling Wheel archive, it is possible to develop a knowledge repository to support a digital platform

The main objectives of this project are the design and development of a:

1. Ontology
2. Knowledge Repository
3. Digital Platform

Objectives

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Research Hypothesis

From the identification of concepts and respective relations over the Foundling Wheel archive, it is possible to develop a knowledge repository to support a digital platform

The main objectives of this project are the design and development of a:

1. Ontology
2. Knowledge Repository
3. Digital Platform

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Objectives

Research Hypothesis

From the identification of concepts and respective relations over the Foundling Wheel archive, it is possible to develop a knowledge repository to support a digital platform

The main objectives of this project are the design and development of a:

1. Ontology
2. Knowledge Repository
3. Digital Platform

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Objectives

Research Hypothesis

From the identification of concepts and respective relations over the Foundling Wheel archive, it is possible to develop a knowledge repository to support a digital platform

The main objectives of this project are the design and development of a:

1. Ontology
2. Knowledge Repository
3. Digital Platform

Documents

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Figure: Book Cover

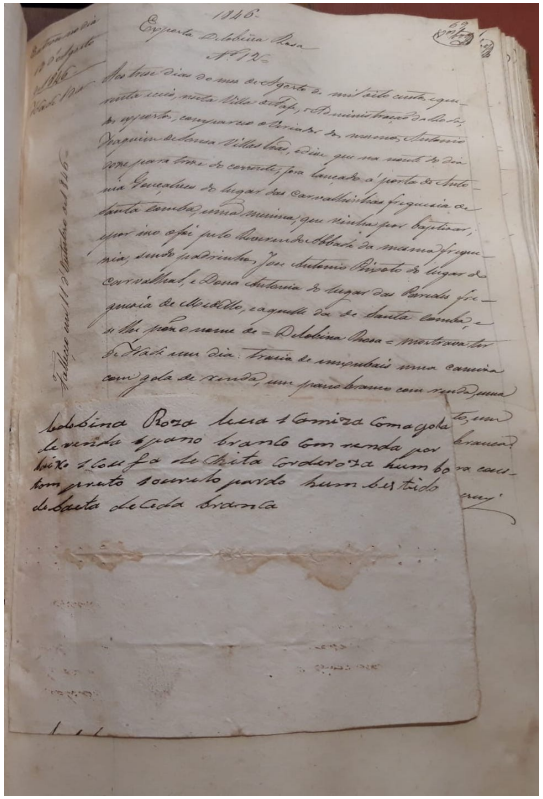


Figure: Document and Signal

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Documents

Classification of documents according to their main topic:

- Entrance Register
- Departure Register
- Medical Records
- Vaccination Records
- Register of Expenses
- Record of Nurses
- Payment to Nurses
- Foundlings given to Nurses
- Foundlings given to their Mothers
- Record of Letters
- Movement of Foundlings

Documents

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Main issues found in these documents:

- Linguist issues,
- Degradation issues

Details on the resulting digital representation:

- Documents provided in a Microsoft Word format,
- Additional punctuation to mark degradation issues,
- Key-Value pairs to describe each book

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Documents

Main issues found in these documents:

- Linguist issues,
- Degradation issues

Details on the resulting digital representation:

- Documents provided in a Microsoft Word format,
- Additional punctuation to mark degradation issues,
- Key-Value pairs to describe each book

Architecture

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

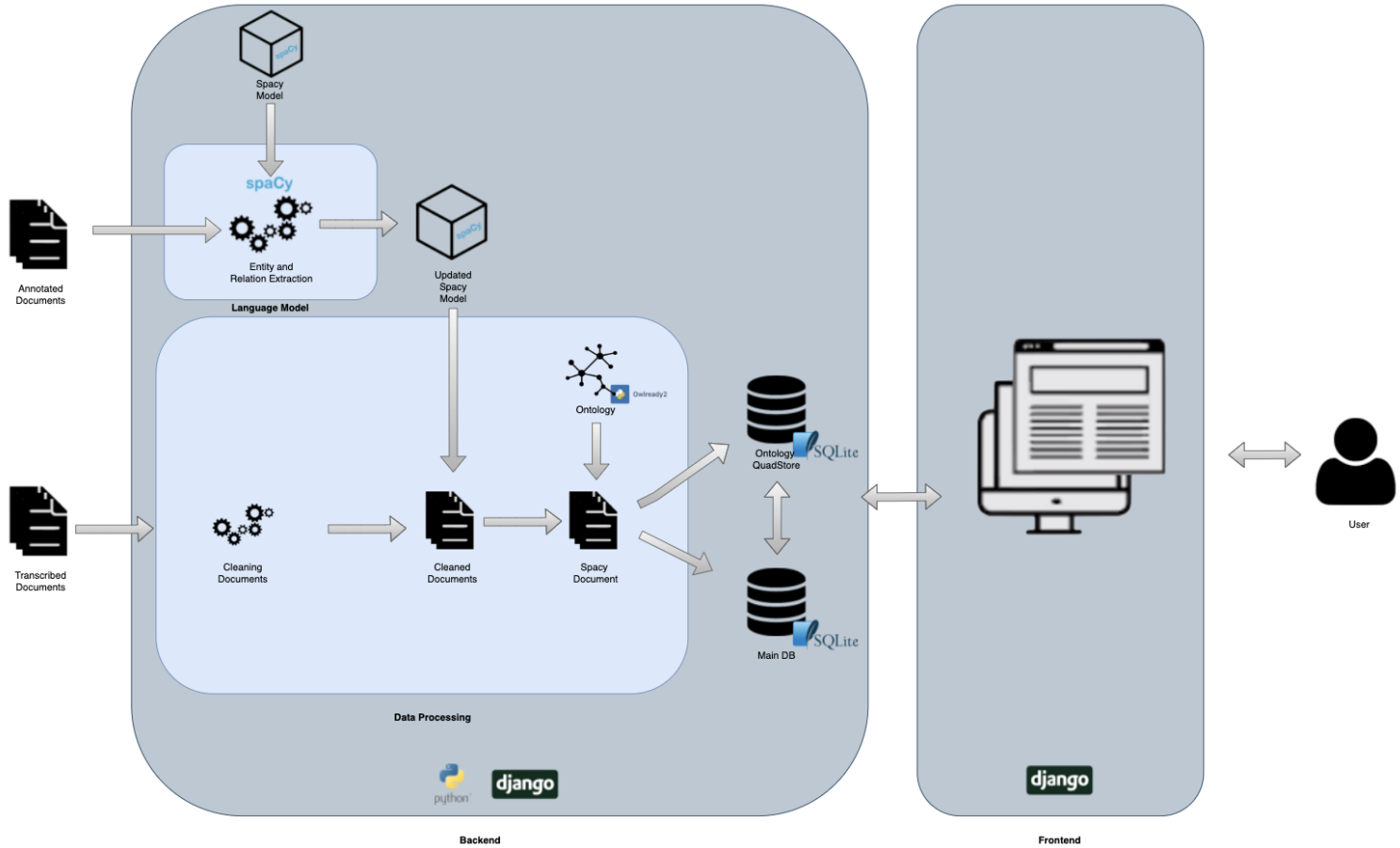


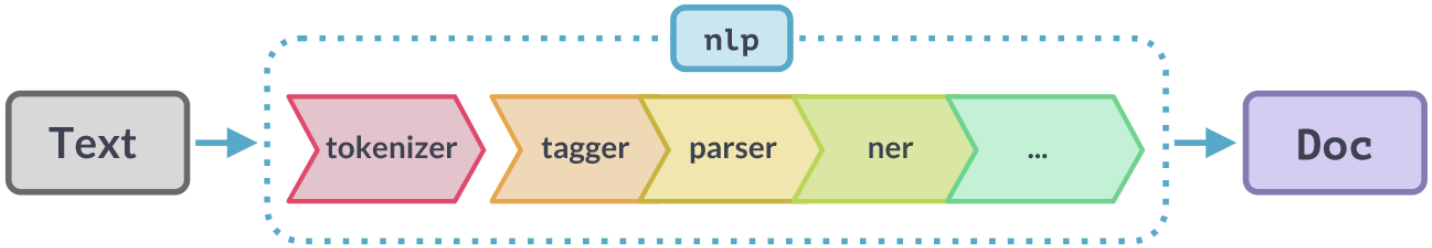
Figure: General Architecture

Architecture

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- Adaptable Natural Language Model. These linguistic models perform standard NLP tasks.

- Automate the annotation process, by processing a document through a pipeline.



Credit: *Spacy Documentation*

Figure: Spacy Processing Pipeline

Architecture

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

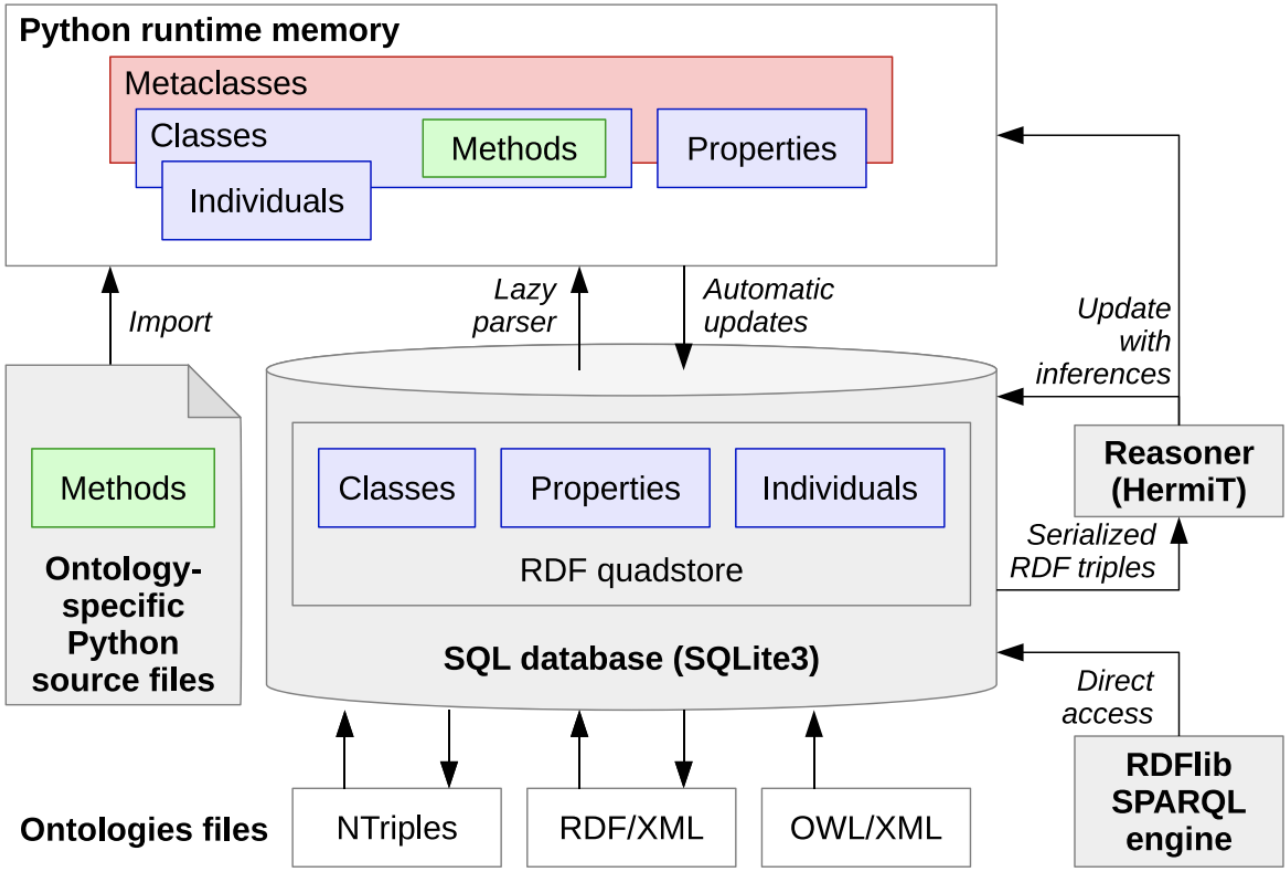


Figure: OWLReady Architecture

Architecture

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

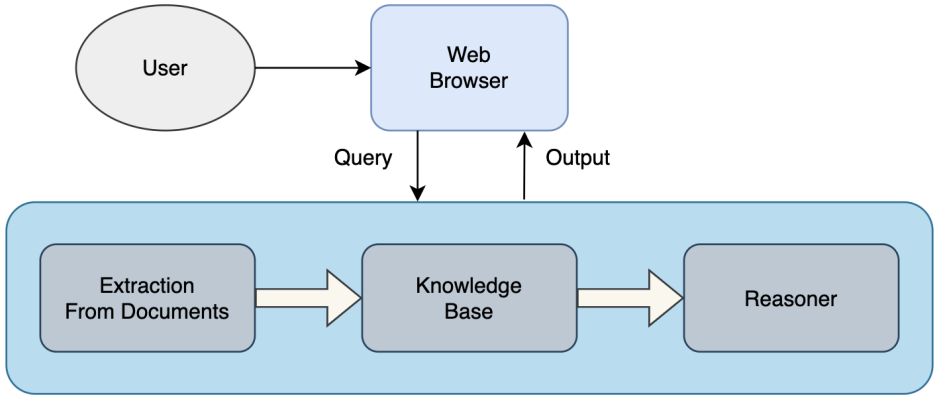


Figure: Repository Interaction

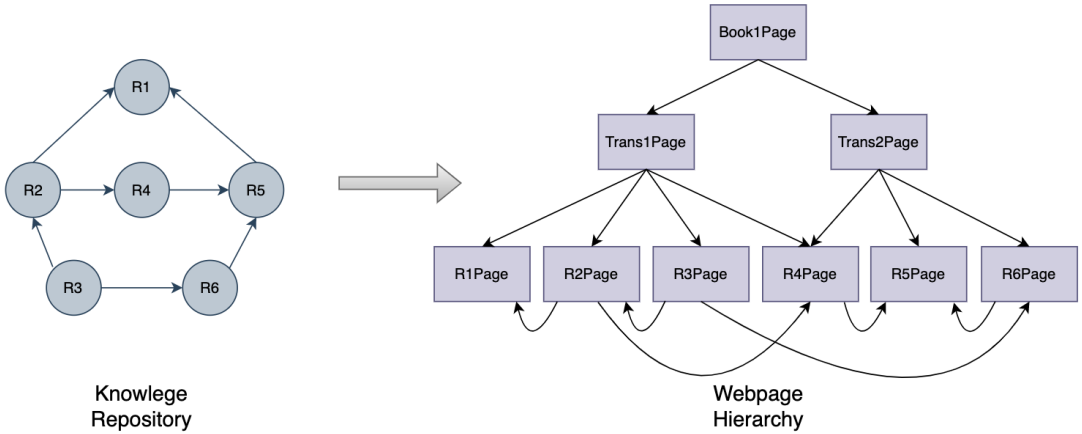


Figure: Repository Rendering

Models and Persistence

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

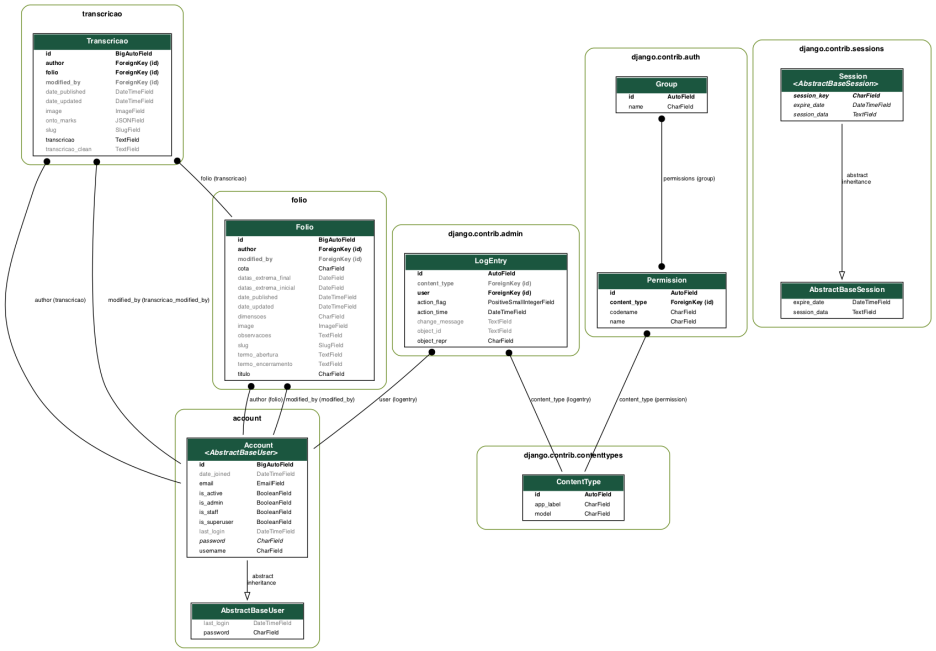


Figure: Database ERD

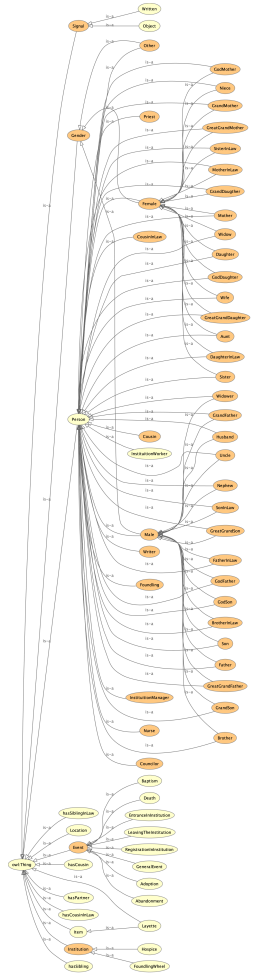


Figure: Ontology Hierarchy

Entity Extraction

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

- Spacy provide a NER trainable component, responsible for labelling sequences of text.

- The goal is to extract from each document a set of named entities, accordingly to the ontology concepts and attributes.

- The standard Portuguese model is retrained and permanently stored.

Figure: NER Identified by the Model

Relation Extraction

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

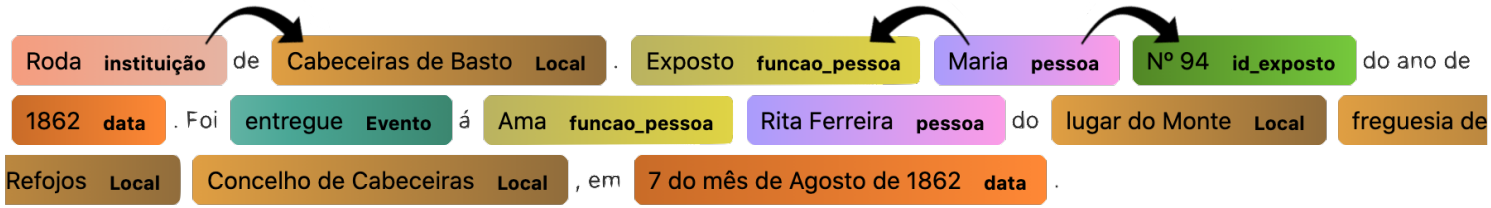


Figure: NER Identified by the Model

- The goal is to extract meaningful relations between named entities.

Relation Extraction

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Figure: Entities Identified

```
pessoa_local = [
            {'ENT_TYPE' : 'pessoa'},
            {"POS": "PUNCT", "OP": "?"},
            {"POS": "ADP", "OP": "?"},
            {'ENT_TYPE' : 'Local'}
              ]
```

Figure: Matcher Pattern

Relation Extraction

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

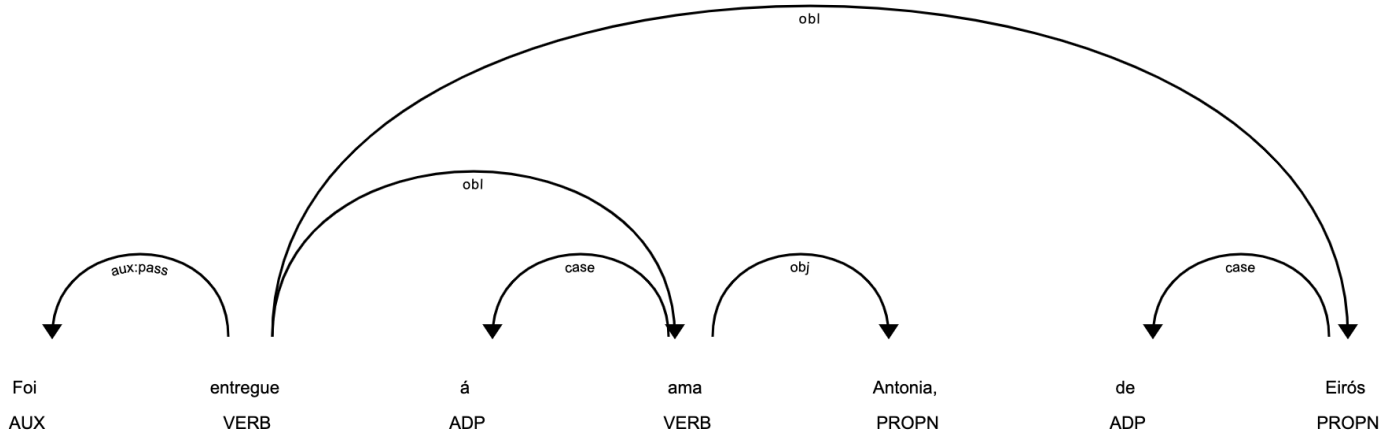


Figure: Dependency Example

```
evento_pessoa = [
    {
        "RIGHT_ID": "evento",
        "RIGHT_ATTRS": {"ENT_TYPE": "Evento"}
    },
    {
        "LEFT_ID": "evento",
        "REL_OP": ">>",
        "RIGHT_ID": "participa",
        "RIGHT_ATTRS": {"ENT_TYPE": "pessoa"}
    }]
```

Figure: Dependency Matcher Pattern

Pipeline Integration

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Transcription processing pipeline, responsible for establishing the connection between the independent repositories.

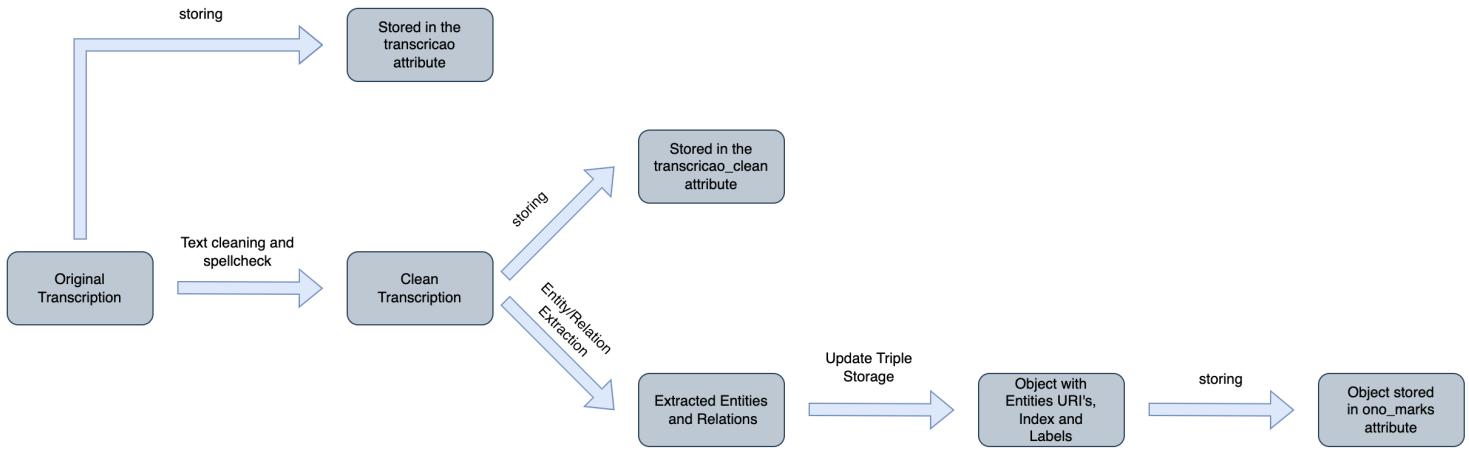


Figure: Transcription Processing Pipeline

Navigation Example

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

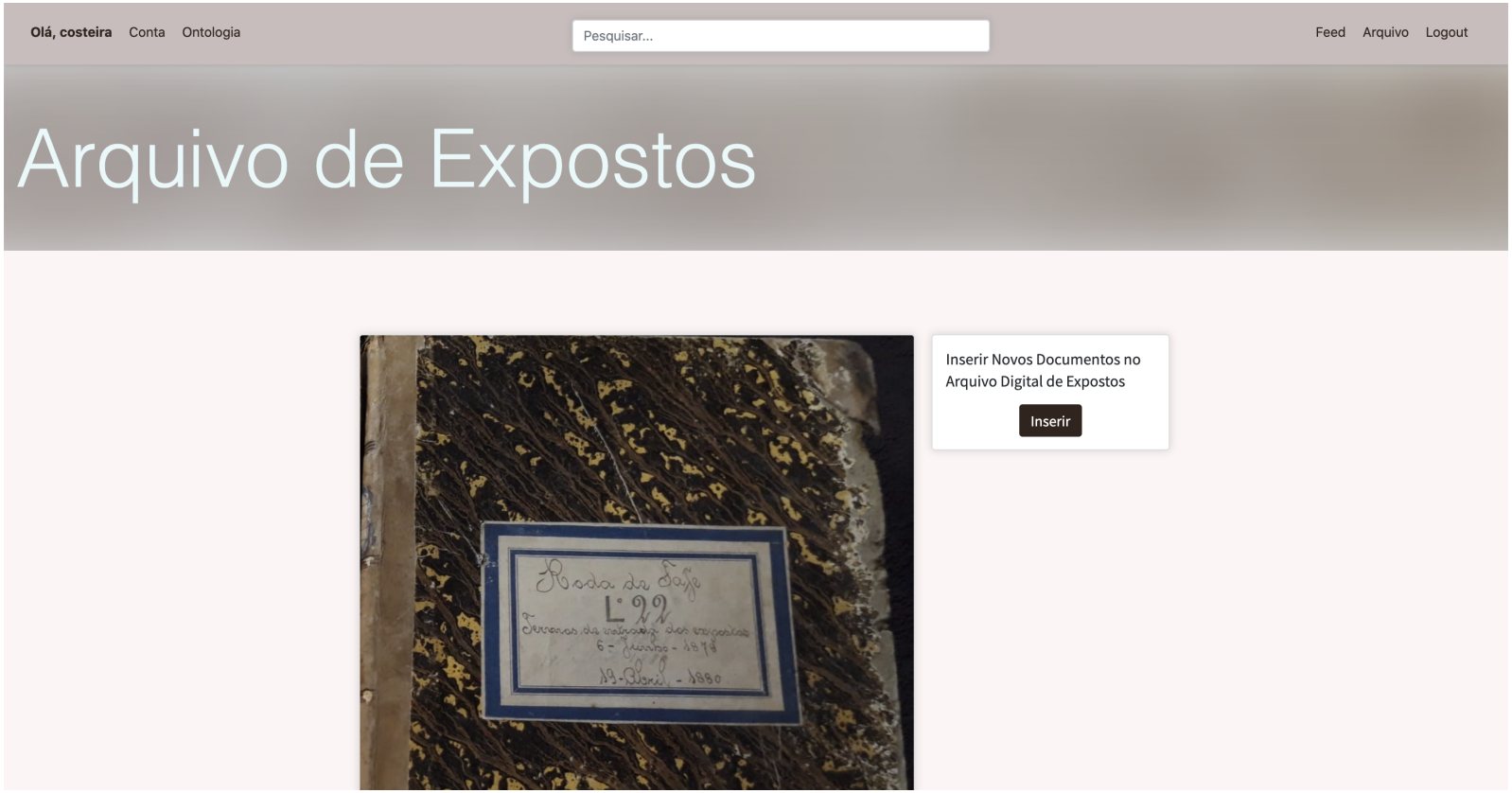


Figure: Feed Page

Navigation Example

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

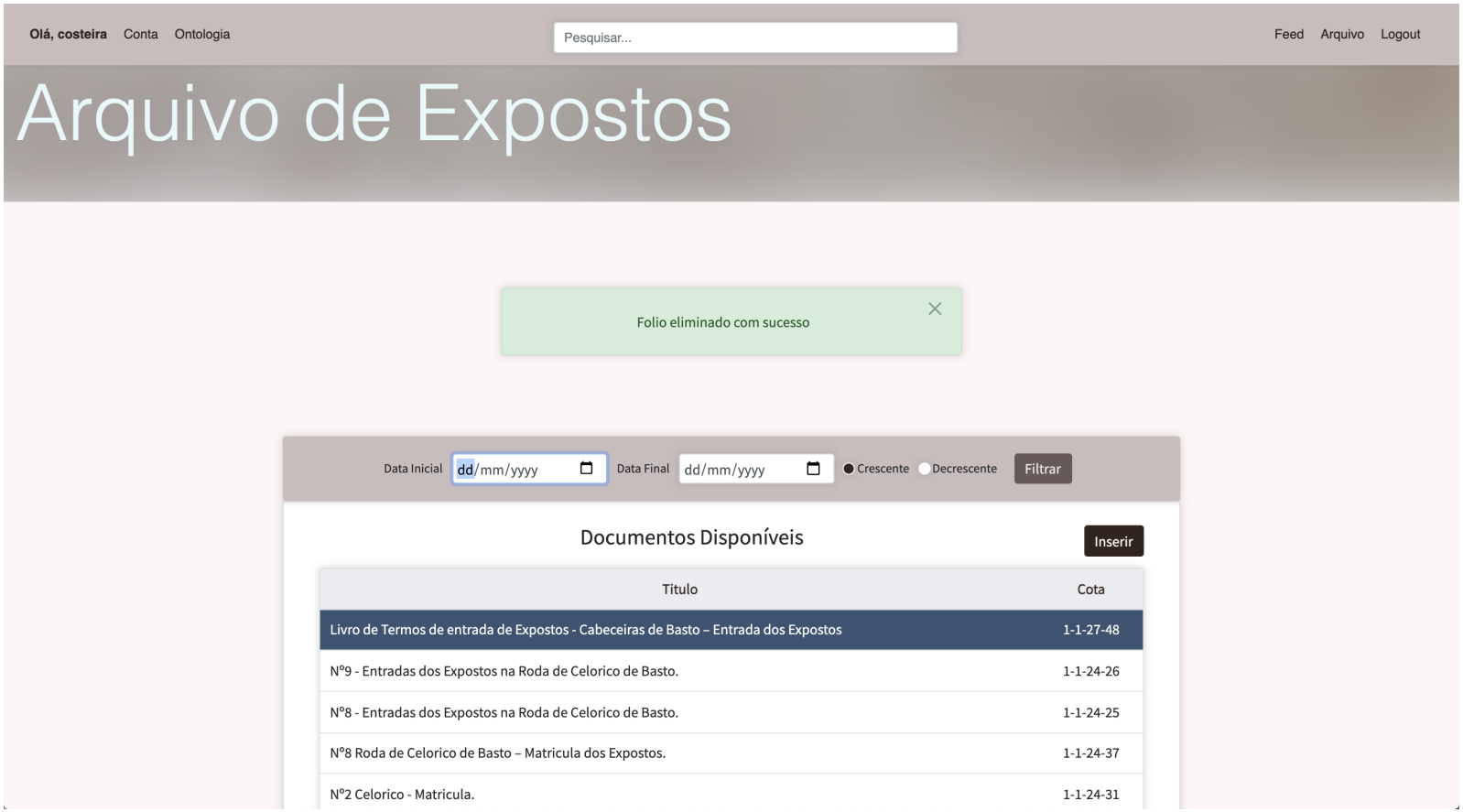


Figure: Book Table

Navigation Example

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Figure: Transcription Page

Navigation Example

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

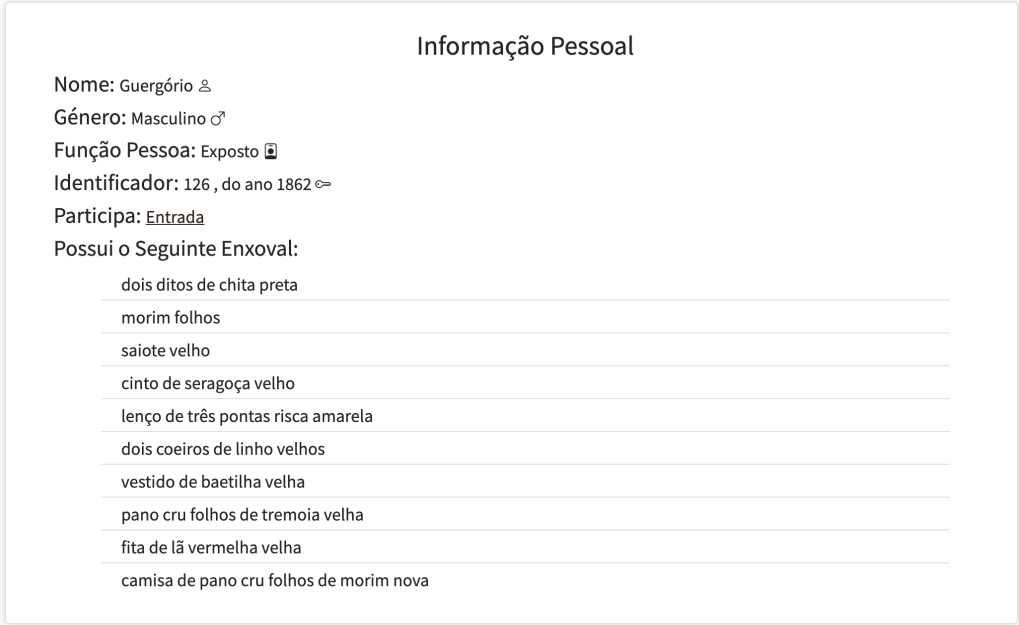


Figure: Personal Information

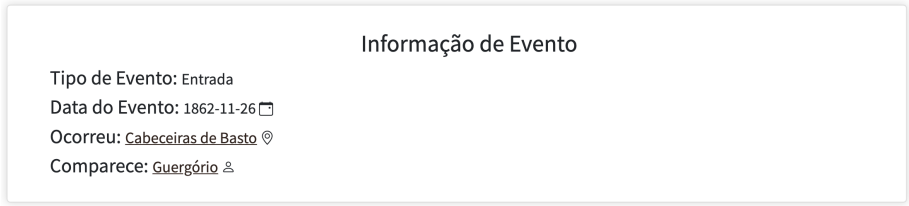


Figure: Event Card

Navigation Example

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

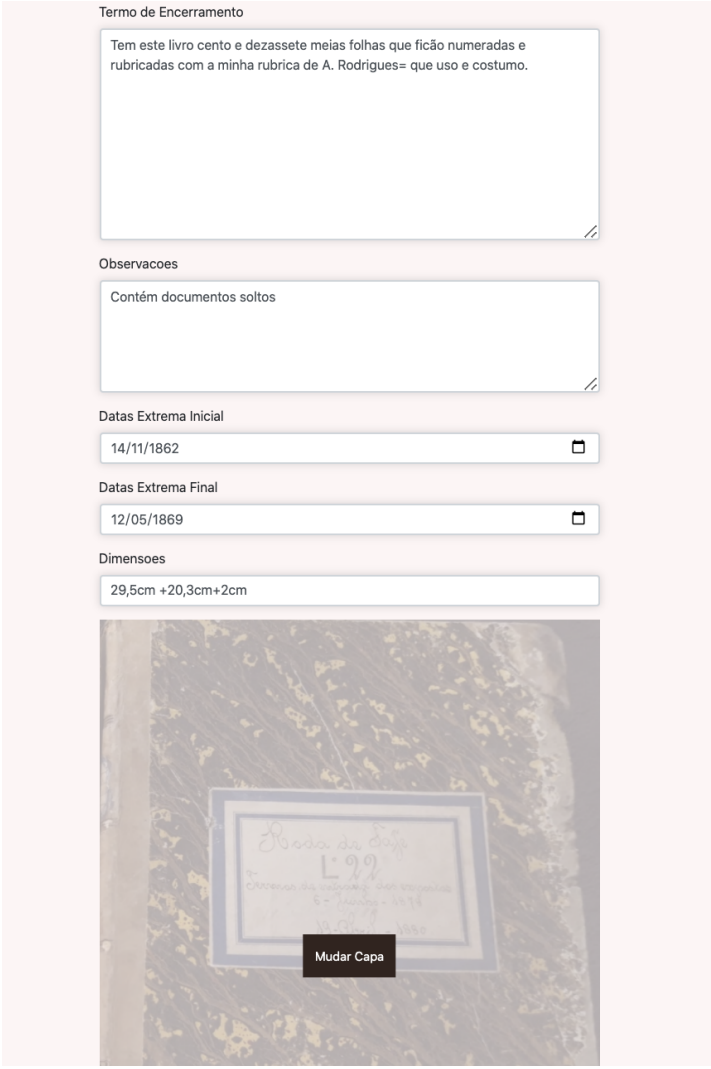


Figure: Update Book Form

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Conclusion

- Research Hypothesis verified, by the development of a knowledge repository that supports a digital platform.

- The proposed solution automates the extraction of knowledge from documents, resorting to the integration of Natural Language components.

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Conclusion

As future Work, a set of improvements were identified:

- Expand the knowledge repository,
- Incorporate an annotation tool in the application,
- SPARQL engine for an advanced user base,
- Integration with similar projects

Contextualization
Motivation
Objectives
Documents
Architecture
Implementation
Models and Persistence
Entity Extraction
Relation Extraction
Pipeline Integration
Example
Conclusion

Exploration of documents concerning Foundlings in Fafe along XIX Century

João Costeira Faria Gomes

Universidade do Minho

2022