### Semantic search

Creating and querying semantic index

#### Jan Kouřil

Natural Language Processing Group Faculty of Information Technology Brno University of Technology

May 16, 2013

Supervisor: Doc. RNDr. Pavel Smrž, Ph.D.



### Outline

- 1
- Motivation and Objectives
- Introduction
- The Decipher project
- Goals of work
- 2
- Semantic index
- Index structure
- Query language
- 3
- **Features**
- Autocomplete
- Classification
- User feedback
- Facets
- Similarity search
- 4
- **Future Work**
- Payloads
- Knowledge-based co-reference





### Outline

- 1
- Motivation and Objectives
- Introduction
- The Decipher project
- Goals of work
- 2

#### Semantic inc

- Index structure
- Query language
- 3

#### Features

- Autocomplete
- Classification
- I Isar faadhad
- Facets
- Similarity search
- 4

#### Future Worl

- Payload
- Knowledge-based co-reference





### Motivation

- Search accuracy can be improved by understanding searcher's intent
- Contextual meaning of lexical units can be resolved easier for specific domain
- Knowledge base for any domain may be used by many different users
- Integrating user feedback is a way to improve search
- Rich query language may help a lot



## Techniques used

- Named entity recognition
- Co-reference resolution
- Finite automata for lexical purposes
- Classification algorithms for document analysis
- Document indexing



## Decipher overview

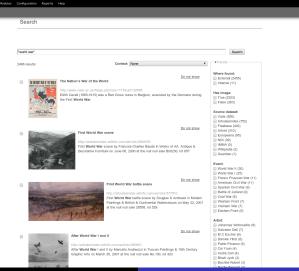
- Digital Environment for Cultural Interfaces
- Promoting Heritage
- Education and Research
- Storyscope front-end
- SEC back-end



6/29

Future Work

### Storyscope



## Semantic Enrichment Component

- Query language
- Faceted search
- Event detection
- Handling index
- Content classification



### Dossier creation

- Search artworks with query language
- Add artworks to dossier
- Extract events and form plots
- Create narratives
- Publish dossier



J. Kouřil

### Outline



#### Motivation and Objectives

- Introduction
- The Decipher project
- Goals of work



#### Semantic index

- Index structure
- Query language



#### Features

- Autocomplete
- Classification
- User feedback
- Facets
- Similarity search



#### Future Work

- Payload
- Knowledge-based co-reference





### Elasticsearch

- Based on Apache Lucene
- RestFul server
- Tools for analysis (tokenizers, stemmers)
- Key-value store
- Multicast cluster



## Indexed types

- Dossiers
- Object stories
- Artworks
- References
- Events
- Plot elements
- Plots



## Example 1

Mona Lisa; dataset: wikipedia; intended_audience: youths			
	Context:	None	

Do not show





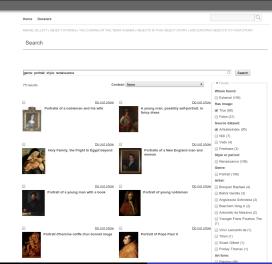
http://simple.wikipedia.org/wiki/Mona\_Lisa

Mona Lisa (also known as La Gioconda or La Joconde) is a 16th-century portrait painted in oil by Leonardo da Vinci during the Renaissance in Florence, Italy. As of 2011, the work is owned by the Government of France and is shown at the Musée du Louvre in Paris under the title Portrait of Lisa Gherardini, wife of Francesco del Giocondo. Many people think Mona Lisa's smile is mysterious. It is so often studied, recognized, and copied that many people say it is the most famous painting in the world. The Louvre says that about 80 percent of its visitors come to see the painting of Mona Lisa.

Vasari, who was Leonardo's first biographer (a person who writes about the life of another person), thought the painting was of a person named Lisa Gherardini. Speculation over the painting's model was solved in 2008 by Dr Armin Schlechter, a manuscript expert. Notes discovered in Heidelberg University Library which were written by Agostino Vespucci, a Florentine city official, reinforced Vasaril's earlier identification of the model. Lisa was the wife of Francesco del Giocondo a wealthy silk merchant, who lived in Florence. Although the sitter has traditionally been identified as Lisa del Giocondo, a lack of definitive evidence has long fueled alternative theories, including Leonardo's mother Caterina in a distant memory as the painting's



# Example 2



## Example 3

person.style: impressionism OR postimpressionism; mentions.place: not in Europe

Context: None

Do not show



#### Paul Gauguin

http://en.wikipedia.org/wiki/Paul\_Gauguin

Gauguin was born in Paris, France, to journalist Clovis Gauguin and Alina Maria Chazal, daughter of the proto-socialist leader Flora Tristan, a feminist precursor whose father was part of an influential Peruvian family. In 1850 [4] the family left Paris for that country, motivated by the political climate of the period. [citation needed] Clovis died on the voyage, leaving eighteenmonth-old Paul, his mother, and sister, to fend for themselves. They fived for four years in Lima with Paul's uncle and his family. The imagery of Peru would later influence Gauguin in his art. It

was in Lima that Gauguin encountered his first art. His mother admired Pre-Columbian pottery, collecting Inca pots that some colonists dismissed as barbaric. "One of Gauguin's few early memories of his mother was of her wearing the traditional costume of Lima, one eye peeping from behind her manteau, the mysterious one-eye veil that all women in Lima went out in. [...] He was always drawn to women with a 'traditional' look. This must have been the first of the colourful fermale costumes that were to haunt his imagination."[5] At the age of seven, Gauguin and his family returned to France, moving to Orléans to live with his grandfather. The Gauguins came originally from the area and were market gardeners and greengrocers; gauguin means 'walnut-grower'. His father had broken with family tradition to become a journalist in Paris. Gauguin soon learned French, though his first and preferred language remained Peruvian Spanish, and he excelled in his studies. After attending a



## **Implementation**

- Implemented in Python
- BoolQuery with list of constraints
- Named Entity Extraction
- Information hierarchy expansion
- Query expansion



### Outline

- Introduction

- - Index structure

  - **Features** 
    - Autocomplete
    - Classification
    - User feedback
    - Facets
    - Similarity search
- - Knowledge-based co-reference





### **Features**

- Query autocomplete
- Facets
- Similarity search
- User feedback
- Classification



## Autocomplete

- Artist names
- Artworks
- Event activities
- Genres
- Artistic styles
- Locations
- Event activities
- Others...



## Autocomplete implementation

- Wildcard queries
- Index lookup
- Geonames database
- Autocomplete with constraints



### Classification

- Choose categories to classify
- Select features
- Choose algorithm
- Take facet



### User feedback

- Positive and negative feedback
- User-specific
- Improve average ranking



Autocomplete
Classification
User feedback
Facets
Similarity search

### **Facets**

- From datasets
- Classified facets
- Events
- Dynamic facets



## **Techniques**

- MoreLikeThis
- MoreLikeThese
- TF-IDF query terms
- Query constraints
- LDA



Autocomplete Classification User feedback Facets Similarity search

## LDA experiments

- The most similar topic
- Overall topic similarity
- The most overlapping topics
- Combined measures
- Synonymy replacement



### Outline



#### Motivation and Objectives

- Introduction
- The Decipher project
- Goals of work



#### Semantic index

- Index structure
- Query language



#### Features

- Autocomplete
- Classification
- User feedback
- Facets
- Similarity search



#### Future Work

- Payloads
- Knowledge-based co-reference





## Future development

- Improve metadata querying
- Implement payload queries
- Improve query language
- Knowledge-based co-reference



## Payloads

Position	Word	NER
1	The	Person
2	man	Person
3	went	Activity
4	to	Activity
5	the	Location
6	store	Location



## Knowledge-based co-reference

#### Examples

- Later, he moved to the south of France and was influenced by the strong sunlight he found there. His work grew brighter in color...
- David and The Last Judgement are perhaps the best known artworks from Italian renaissance artist Michelangelo. The sculpture is located in Galleria dell'Accademia, Florence.

#### Extractions

- An artist (Vincent van Gogh) was influenced by an event (moving to south France)
- Sculpture of David is located in Florence.

