FONDAZIONE BRUNO KESSLER

DKM DATA & KNOWLEDGE MANAGEMENT

Collaborative NLP-aided ontology modelling

Chiara Ghidini
ghidini@fbk.eu

Marco Rospocher
rospocher@fbk.eu

International Winter School on Language and Data/Knowledge Technologies
TrentoRISE – Trento, 24th February 2012

FONDAZIONE
BRUNO KESSLER

Part I

ONTOLOGIES & ONTOLOGY MODELLING

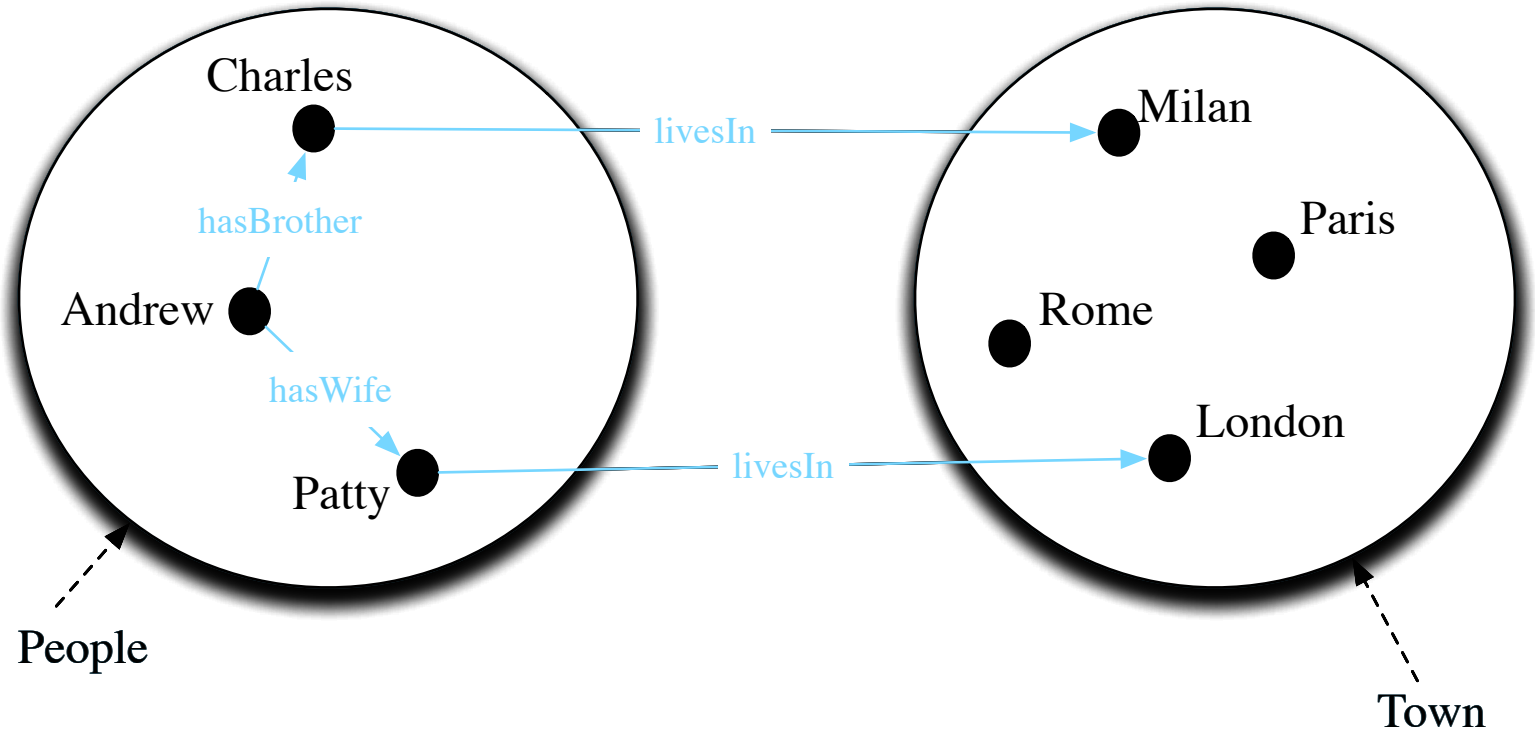
FONDAZIONE BRUNO KESSLER

What is an ontology?

- Many definitions of an ontology in literature;
- Here we refer to an ontology as a "formal specifications of the terms in the domain and relations among them" (*)
- Ontologies contain a formal explicit description of:
  - Concepts (aka classes)
  - Relations (aka roles)
  - Individuals (aka instances)
- Classes (and relations) can be ordered in taxonomies using the subclass relation

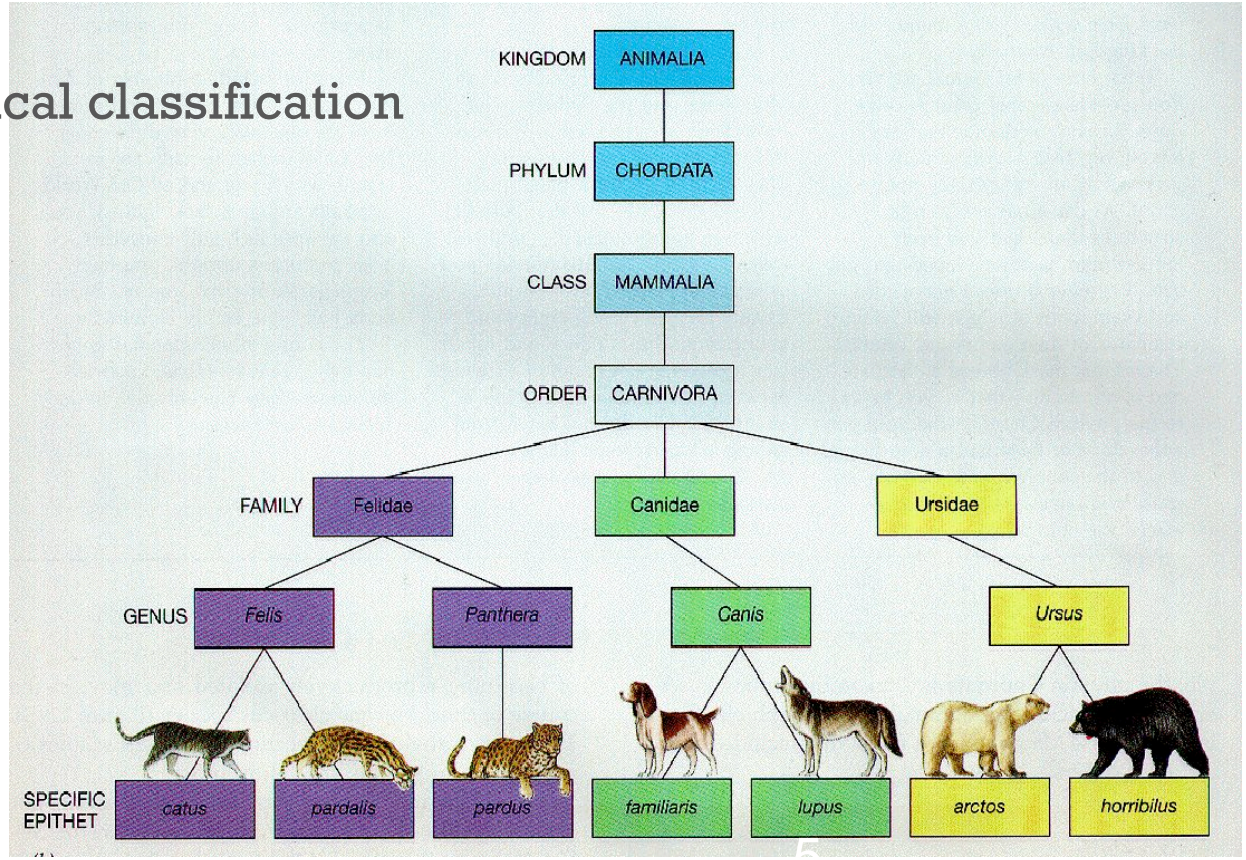
(*) [Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5: 199-220.]

In a picture


Charles
livesIn
Milan
hasBrother
Paris
Andrew
Rome
hasWife
London
Patty
livesIn
People
Town

Taxonomies

- Classes (and relations) can be ordered in taxonomies using the subclass relation
- Example: biological classification of species
- Same for roles



FONDAZIONE BRUNO KESSLER

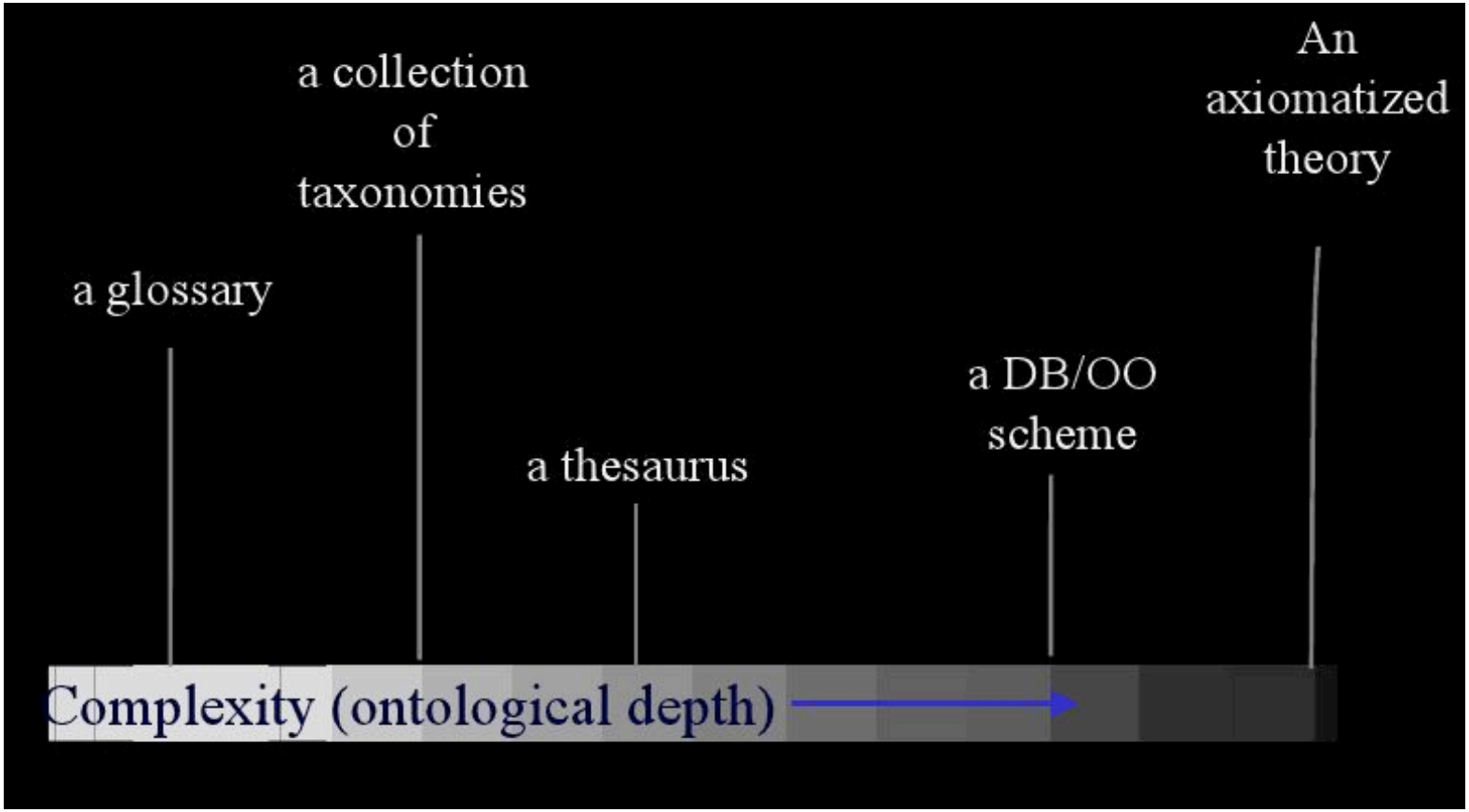
Axioms

- Concepts can be formally described through axioms

- *A Pizza Margherita is a pizza which has both tomato topping and mozzarella topping*

$$PizzaMargherita \sqsubseteq Pizza$$
$$PizzaMargherita \sqsubseteq \exists hasTopping.TomatoTopping$$
$$PizzaMargherita \sqsubseteq \exists hasTopping.MozzarellaTopping$$

Different types of Ontologies



Slide taken from “Ontology-Driven Conceptual Modelling” A tutorial by Nicola Guarino.

FONDAZIONE
BRUNO KESSLER

Why to develop an ontology?

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

FONDAZIONE
BRUNO KESSLER

Examples of ontologies

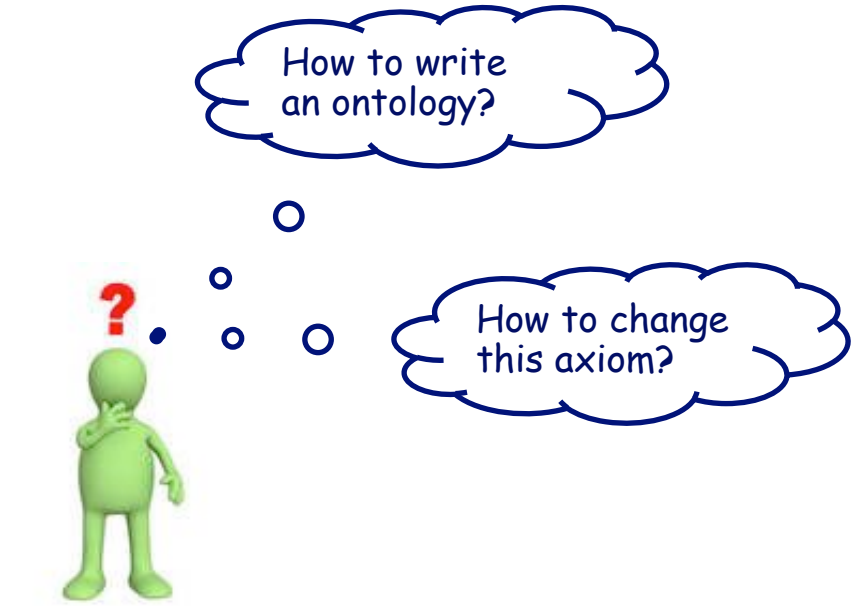
- Large taxonomies categorizing Web sites (such as on Yahoo!)
- Medical Ontologies (such as SNOMED) to annotate documents and share information
- Categorizations of products for sale and their features (such as on Amazon.com, but also smaller enterprises).

- Therefore……

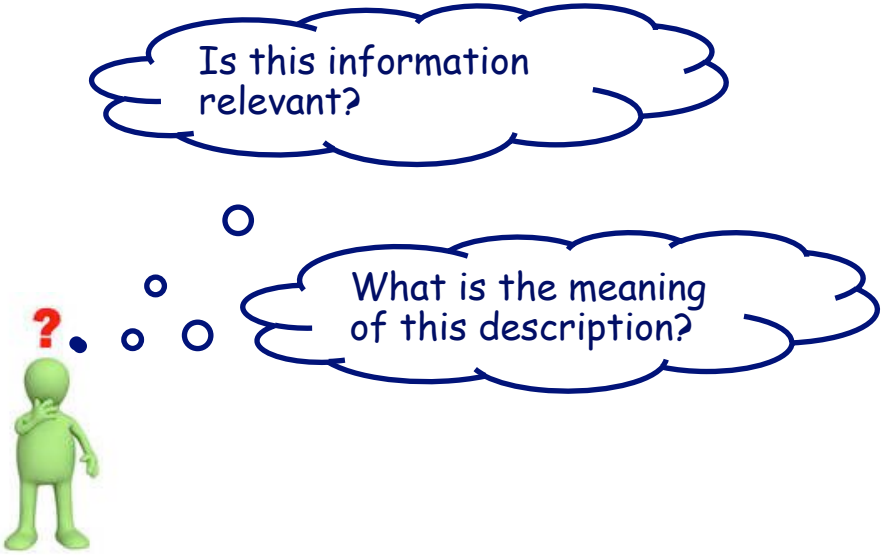
The development of ontologies is moving from the realm of research labs to the "desktop of domain experts"

Problems in ontology modeling

1. Modelling is a collaborative activity



Domain expert



Knowledge engineer

Problems in ontology modeling

2. Modelling is a time-consuming and error-prone activity, and often needs parsing of a large quantity of material.

FONDAZIONE
BRUNO KESSLER

Our contribution

Our Contribution to solve those problems

1. Framework for the collaborative modeling of ontologies using wikis
2. Automatic extraction of key-phrases for ontology modelling

FONDAZIONE
BRUNO KESSLER

Part II

COLLABORATIVE FRAMEWORK FOR ONTOLOGY MODELING

FONDAZIONE BRUNO KESSLER

Why a wiki-based conceptual modeling tool?

- Wikis support **collaborative** editing;
- Users are quite **familiar** with viewing/editing wiki content (e.g. Wikipedia);
- Only a **web-browser** is required on the client side;
- Wikis provide a **shared knowledge repository** accessible by users spread all over the world;
- Wikis can provide a **uniform tool/interface** for the specification of different model types (e.g. ontologies, processes, …);

An architecture for collaborative conceptual modeling in wikis

1. One element ⟺ One page
   - each element of the model is represented by a page in the wiki;

Concept “Mountain” ⟺

Mountain

A **mountain** is a large landform that stretches above the surrounding land in a limited area usually in the form of a peak. A mountain is generally steeper than a **hill**.

The highest mountain on earth is the Mount Everest

FONDAZIONE BRUNO KESSLER

An architecture for collaborative conceptual modeling in wikis

2. Unstructured and structured descriptions
   - each page contains both structured and unstructured content;

Mountain

A **mountain** is a large landform that stretches above the surrounding land in a limited area usually in the form of a peak. A mountain is generally steeper than a **hill**.

The highest mountain on earth is the Mount Everest

(unstructured content)

$$\sqsubseteq Landform$$

$$\sqsubseteq \neg Hill \sqcap \neg Plain$$

$$\sqsubseteq \forall madeOf(Earth \sqcup Rock)$$

$$\sqsubseteq \exists height. \geq_{2500}$$

$$Mountain(Mt.Everest)$$

$$Mountain(Mt.Kilimanjaro)$$

(structured content)

An architecture for collaborative conceptual modeling in wikis

3. Different views to access the model:
   - different views to support different modeling actors;

Mountain

A **mountain** is a large landform
stretches above the surrounding
a limited area usually in the form
peak. A mountain is generally st
than a **hill**.

The highest mountain on earth is the Mount Everest

(unstructured view)

Mountain

is a	landform
different from	hill, plain
made of	earth
made of	rock
height	at least 2,500m
samples	Mt. Everest Mt. Kilimanjaro

(semi - structured view)

Mountain

$\sqsubseteq Landform$

$\sqsubseteq \neg Hill \sqcap \neg Plain$

$\sqsubseteq \forall madeOf(Earth \sqcup Rock)$

$\sqsubseteq \exists height. \geq_{2500}$

$Mountain(Mt.Everest)$

$Mountain(Mt.Kilimanjaro)$

(fully - structured view)

An architecture for collaborative conceptual modeling

- Alignment between the different views

Mountain

A **mountain** is a large landform that stretches above the surrounding land in a limited area usually in the form of a peak. A mountain is generally steeper than a **hill**.

The highest mountain on earth is the Mount Everest

(unstructured view)

is a	landform
different from	hill, plain
made of	earth
made of	rock
height	at least 2,500m
samples	Mt. Everest Mt. Kilimanjaro

(semi-structured view)

$\sqsubseteq Landform$

$\sqsubseteq \neg Hill \sqcap \neg Plain$

$\sqsubseteq \forall madeOf(Earth \sqcup Rock)$

$\sqsubseteq \exists height. \geq_{2500}$

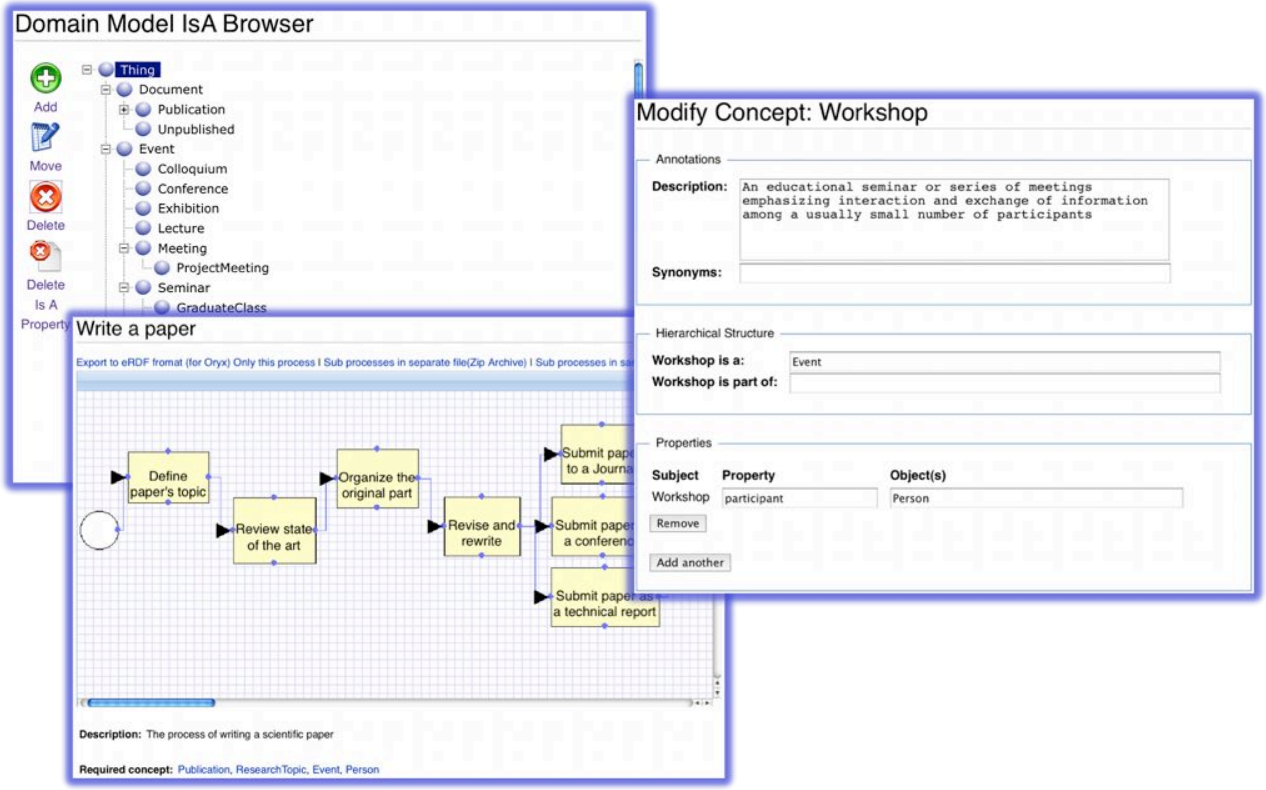
$Mountain(Mt.\_Everest)$

$Mountain(Mt.\_Kilimanjaro)$

(fully structured view)

MoKi: The modeling wiki

MoKi
the Modelling WiKi ---



Web 2.0 tool

Collaborative editing between knowledge experts and knowledge engineers

Term extraction features

Automatic translation from and to OWL and BPMN

Graphical and textual editing

Integrated ontology and process modeling

Support for validation and feedback

Available as open source tool. Demo at moki.fbk.eu

FONDAZIONE
BRUNO KESSLER

Part III

MOKI DEMO

FONDAZIONE
BRUNO KESSLER

Definition of the collaborative framework

Hints on the applicability of the tool also for other conceptual modelling languages (BPMN)

Showcase of results and usages

FONDAZIONE
BRUNO KESSLER

Part IV

AUTOMATIC EXTRACTION OF KEY-PHRASES FOR ONTOLOGY MODELLING

FONDAZIONE BRUNO KESSLER

NLP-aided ontology engineering

- Support ontology modeling by **extracting concepts** characterizing a domain from a **reference text corpus…**
- … actually, by automatically extracting **key-phrases**
- **Key-phrases** are the terms **characterizing** a document or a corpus of documents => candidate relevant **concepts** of the **domain** described by the **corpus**
- Automatic concepts extraction plays an important role in ontology modeling:
  - To boost the ontology **construction/extension** phase
  - To "**validate**" an ontology against a domain corpus

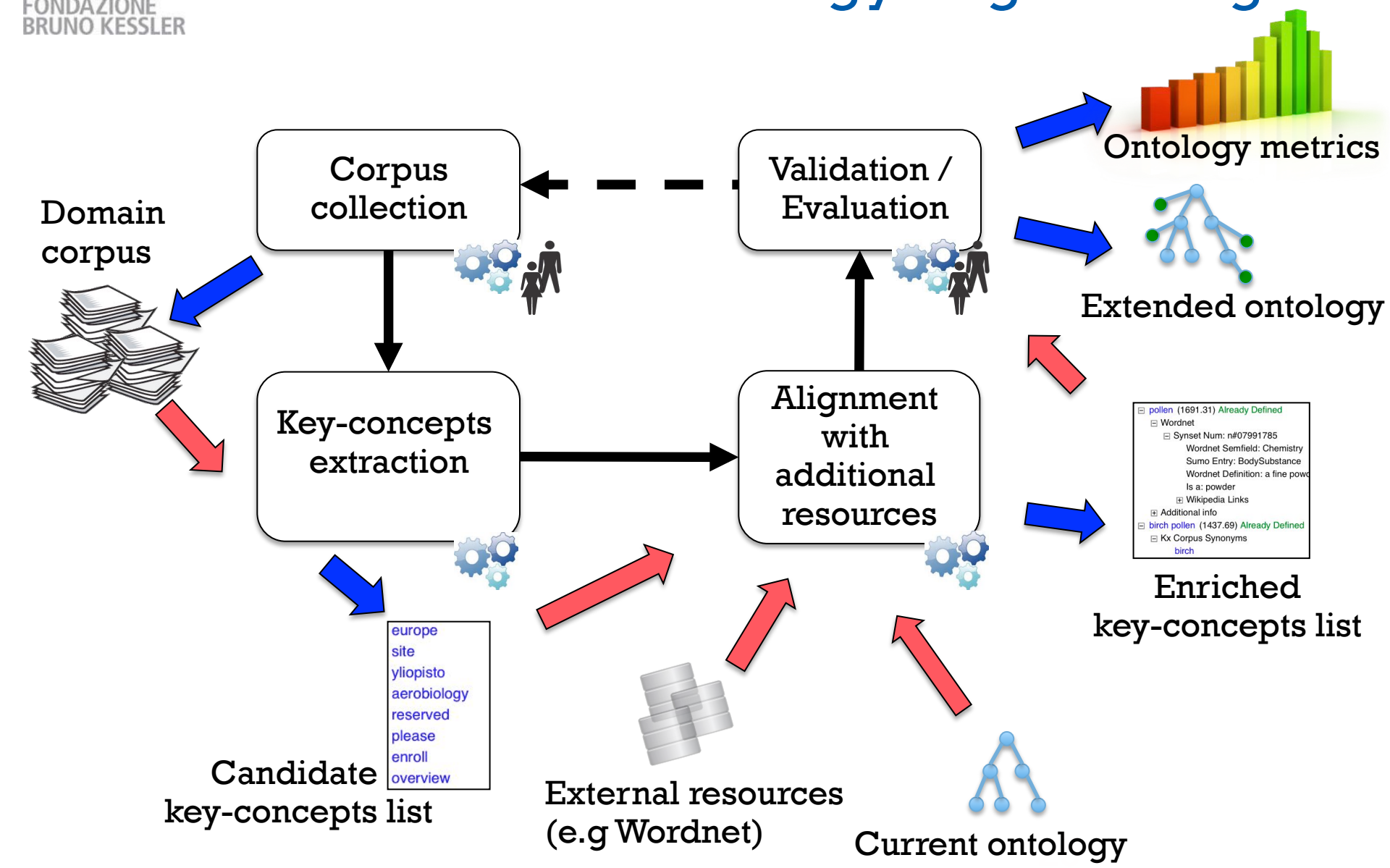
FONDAZIONE
BRUNO KESSLER

An NLP-aided ontology engineering framework

- A framework for supporting ontology **building/evaluation** by automatic **concept extraction** from a reference text corpus

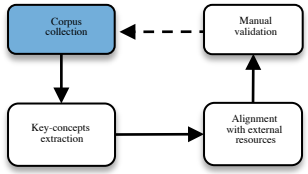
- A fully-working and publicly available **implementation** of the proposed framework in **MoKi**

NLP-aided ontology engineering



FONDAZIONE BRUNO KESSLER

Corpus Selection

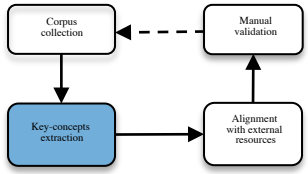


- The corpus can be **manually** or **automatically** selected (e.g. crawling web pages).

- Corpus could consist of:
  - (large) **collection** of documents
    - e.g. pollen bulletins crawled on-line
  - A **single** big document
    - e.g. the BPMN specification.

FONDAZIONE BRUNO KESSLER

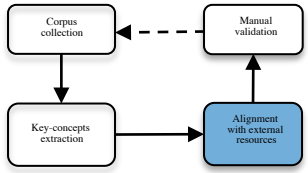
Key-concept extraction



- Performed by **KX** (**K**eyphrase e**X**traction) tool.
  - exploits **linguistic** information and **statistical** measures to select a list of **weighted keywords** from documents;
  - handles **multi-words**;
  - flexible **parameters** configuration;
  - easily adaptable to **new languages**;
  - ranked 2nd (out of 20) at SemEval2010, task on "*Automatic Keyphrase Extraction from Scientific Articles*".

FONDAZIONE BRUNO KESSLER

Alignment with additional resources

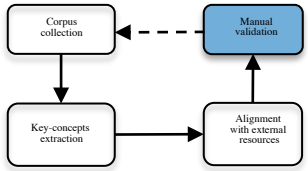


- Extracted key-concepts **aligned** and **enriched** with additional resources:
  - **WordNet** (& WN domains): synonyms, definitions, SUMO labels;
  - **Wikipedia**: link to the Wikipedia page corresponding to the term (exploiting BabelNet);
  - Other external resources (e.g. dictionary).

- Enriched key-concepts list **matched against** the ontology, to detect already defined key-concepts.

FONDAZIONE BRUNO KESSLER

Ontology Extension / Evaluation



- Ontology **Extension**:
  - The user **decides** which of the extracted key-concepts to add to the ontology;
  - The additional details provided in the enriched list may **guide the formalization**;
    - e.g. is-a related synsets, definitions, …

- Ontology **Terminological Evaluation**:
  - Automatically computed metrics (variants of IR **precision** and **recall**) support users in determining the terminological coverage of the ontology wrt to the corpus used;

FONDAZIONE BRUNO KESSLER

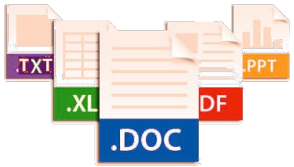
Application Scenarios

- The proposed approach can support several different ontology engineering tasks:
  - **Ontology construction boosting**: building an ontology from scratch;
  - **Ontology extension**: adding new concepts to an existing ontology;
  - **Ontology evaluation**: evaluating terminologically an ontology against a domain corpus;
  - **Ontology ranking**: ranking candidate ontologies wrt a given domain corpus;
  - **Ranking of ontology concepts**: determining which are the domain-wise most relevant concepts defined in an ontology.

FONDAZIONE BRUNO KESSLER

MoKi
the Modelling WiKi ---

- Framework **fully-implemented** in MoKi
- **Publicly available** @ moki.fbk.eu
- Accepts a collection of digital documents in **any popular formats**



- Let's see it in **action**!

FONDAZIONE
BRUNO KESSLER

Part V

MOKI DEMO (CONTINUED)

FONDAZIONE
BRUNO KESSLER

Part VI

PHD CALL ON INFORMATION EXTRACTION FOR ONTOLOGY ENGINEERING

FONDAZIONE
BRUNO KESSLER

Building Quality Ontologies

- **Starting Point:** a collaborative ontology modeling framework supported by NLP techniques
- **Goal**: to support building **rich** and **high quality** ontologies
- **Issue**: current state of the art NLP techniques for information extraction have some limitations wrt ontology modeling:
  - mainly focused on the extraction of **terms**;
  - more suitable to support the construction of **light-weight medium-quality** ontologies;
- **Challenge**: how to appropriately exploit NLP techniques to support the construction of rich and high quality ontologies?

FONDAZIONE BRUNO KESSLER

PhD call on Information Extraction for Ontology Engineering

- **Objective**:

  Investigate how to combine work in **automatic ontology learning** and work in **methodologies and tools** for manual **knowledge engineering** to produce (semi)-automatic services for ontology learning better supporting the construction of **rich and good quality ontologies**.

- Address **key research challenges** in NLP and ontology engineering.

- Strong **algorithmic** and **methodological** aspects, together with **implementation**-oriented tasks.

FONDAZIONE BRUNO KESSLER

MoKi team at FBK

the Modelling WiKi ---

Collaborative modeling of ontologies and processes

Ontological description of processes

Guided domain expert modeling via template

Multi-linguality and eGovernment application

FONDAZIONE BRUNO KESSLER

DKM DATA & KNOWLEDGE MANAGEMENT

Thank You! Questions?

Chiara Ghidini

http://dkm.fbk.eu/ghidini

ghidini@fbk.eu

Marco Rospocher

http://dkm.fbk.eu/rospocher

rospocher@fbk.eu