

Information Extraction

Introduction

What is Information Extraction

The task of Information Extraction is to identify instances of a particular **pre-specified** class of **entities**, **relationships** and **events** in natural language texts, and the extraction of the **relevant properties** (arguments) of the identified entities, relationships or events. The information to be extracted is pre-specified in user-defined structures called templates (or objects), each consisting of a number of slots (or attributes), which are to be instantiated by an IE system as it processes the text.

Taken from: Information Extraction: Past, Present and Future

- Pre-Specified means, that it is known in advance and therefore highly domain-specific (as opposed to Open Information Extraction)

Information Extraction

- This definition established different tasks:

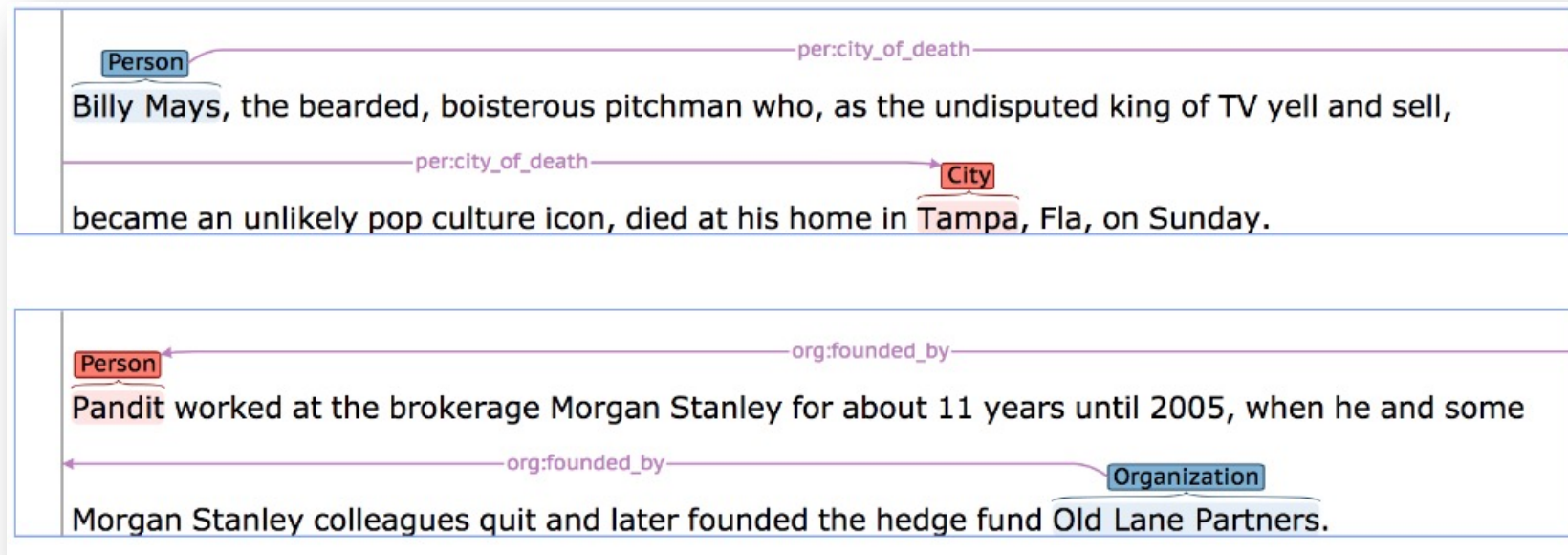
1. Named Entity Recognition

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported **ORG** byF.B.I. Agent **Peter Strzok PERSON** ,
Who Criticized Trump **PERSON** in Texts, Is FiredImagePeter Strzok, a top **F.B.I. GPE** counterintelligence agent who was taken off the special counsel
investigation after his disparaging texts about President **Trump PERSON** were uncovered, was fired. CreditT.J. Kirkpatrick **PERSON** for The New York
TimesBy Adam Goldman **ORG** and Michael S. SchmidtAug **PERSON** . 13 **CARDINAL** , 2018WASHINGTON **CARDINAL** — Peter Strzok
PERSON , the **F.B.I. GPE** senior counterintelligence agent who disparaged President **Trump PERSON** in inflammatory text messages and helped
oversee the **Hillary Clinton PERSON** email and **Russia GPE** investigations, has been fired for violating bureau policies, Mr. **Strzok PERSON** 's lawyer
said **Monday DATE** .Mr. Trump and his allies seized on the texts — exchanged during the **2016 DATE** campaign with a former **F.B.I. GPE** lawyer,
Lisa Page — in **PERSON** assailing the **Russia GPE** investigation as an illegitimate "witch hunt." Mr. **Strzok PERSON** , who rose over **20 years**
DATE at the **F.B.I. GPE** to become one of its most experienced counterintelligence agents, was a key figure in **the early months DATE** of the
inquiry.Along with writing the texts, Mr. **Strzok PERSON** was accused of sending a highly sensitive search warrant to his personal email account.The
F.B.I. GPE had been under immense political pressure by Mr. **Trump PERSON** to dismiss Mr. **Strzok PERSON** , who was removed **last summer**
DATE from the staff of the special counsel, **Robert S. Mueller III PERSON** . The president has repeatedly denounced Mr. **Strzok PERSON** in posts on

Information Extraction

- This definition established different tasks:

2. Relation Extraction



Information Extraction

- This definition established different tasks:

3. Template Filling (or Event Extraction)

| Text | Template |
|---|--|
| <p>EV1 There are no reports of damage or injuries after a small earthquake rattled the Chino Hills area Tuesday morning.</p> | <p><u>EV1</u></p> <ul style="list-style-type: none"> • EVENT: earthquake • DATE: Tuesday morning • TIME: 6:40 a.m • MAGNITUDE: 3.1 • LOCATION: Chino Hills |
| <p>EV1 The 3.1-magnitude quake hit at 6:40 a.m. and was centered about two miles west of Chino Hills.</p> | <p><u>EV1</u></p> <ul style="list-style-type: none"> • EVENT: quake • DATE: Last July • TIME: • MAGNITUDE: 5.4 • LOCATION: |
| <p>EV1 It was felt in several surrounding communities.</p> | |
| <p>EV2 Last July, a 5.4-magnitude quake hit the same area.</p> | |
| <p>EV2 That quake resulted in cracked walls and broken water and gas lines.</p> | |

Information Extraction

- This definition established different tasks:

4. Sentiment Analysis

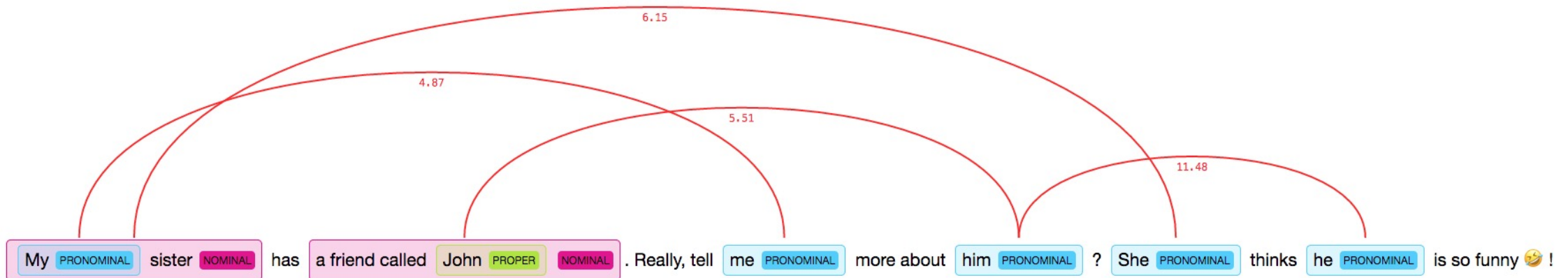
| | |
|--|----------|
| Loves the German bakeries in Sydney. Together with my imported honey it feels like home | Positive |
| @VivaLaLauren Mine is broken too! I miss my sidekick | Negative |
| Finished fixing my twitter...I had to unfollow and follow everyone again | Negative |
| @DinahLady I too, liked the movie! I want to buy the DVD when it comes out | Positive |
| @frugaldougal So sad to hear about @OscarTheCat | Negative |
| @Mofette brilliant! May the fourth be with you #starwarsday #starwars | Positive |
| Good morning thespians a bright and sunny day in UK, Spring at last | Positive |
| @DowneyisDOWNEY Me neither! My laptop's new, has dvd burning/ripping software but I just can't copy the files somehow! | Negative |

Taken from: <https://opendatahub.io/news/2019-09-04/sentiment-analysis-blog.html>

Information Extraction

- This definition established different tasks:

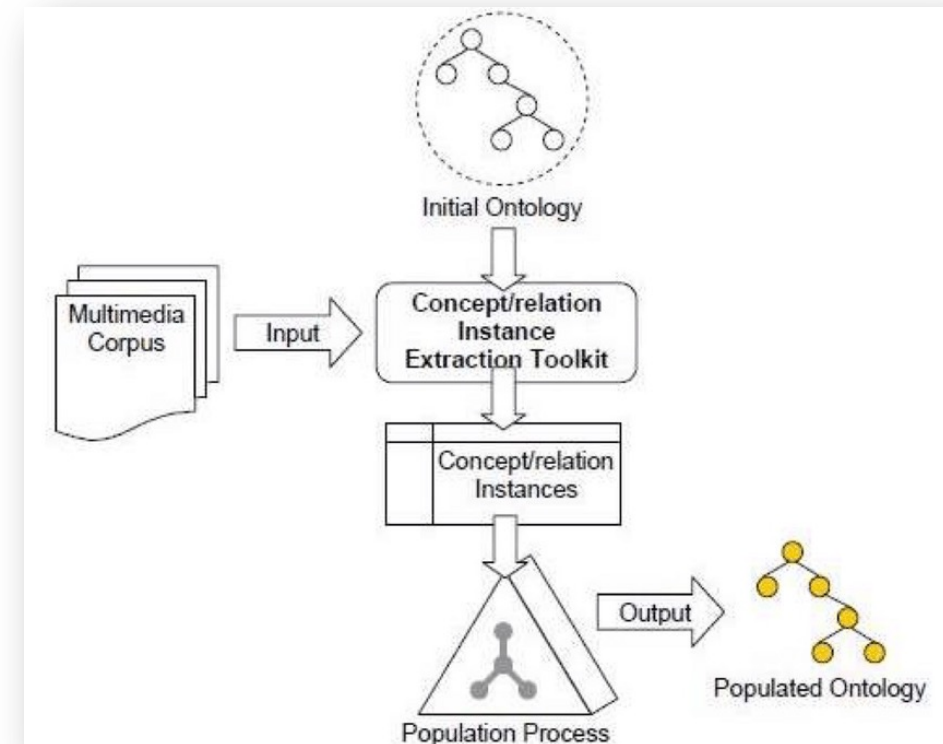
5. Coreference Resolution



Information Extraction

- This definition established different tasks:

6. Ontology Population



https://www.researchgate.net/figure/Ontology-population-process-14_fig1_304412768

Information Extraction

- Commonalities:
 - What these tasks have in common, is that the schema against which we extract (and structure) information takes the form of an **ontology**
 - One way to model this schema is OWL (Web Ontology Language)
 - However I am not going into much detail of OWL

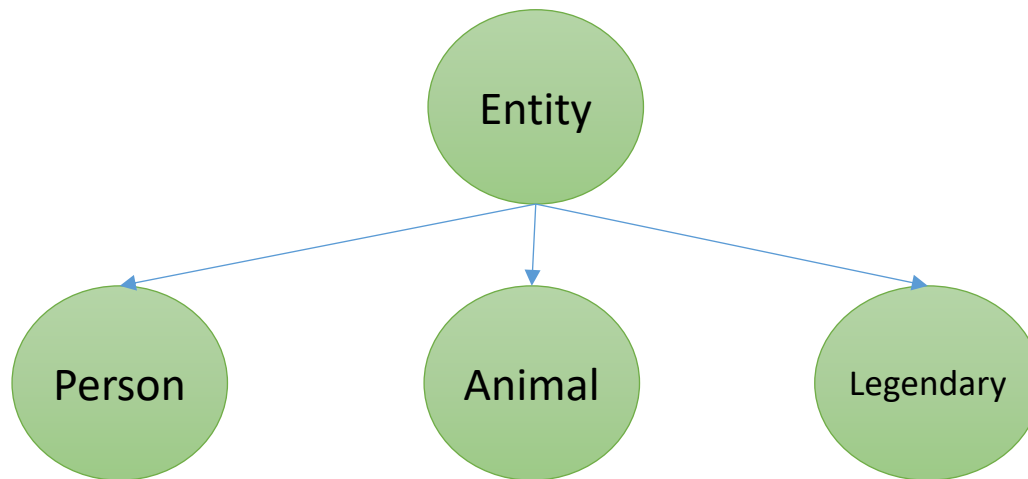


Information Extraction - Ontology

- Our (crude) definition of an ontology requires:
 1. Concepts (or classes), that are arranged in a hierarchy
 2. Attributes of and between the concepts (also called: slots, properties)
 3. Individuals: Concrete examples of the aforementioned concepts

Information Extraction - Ontology

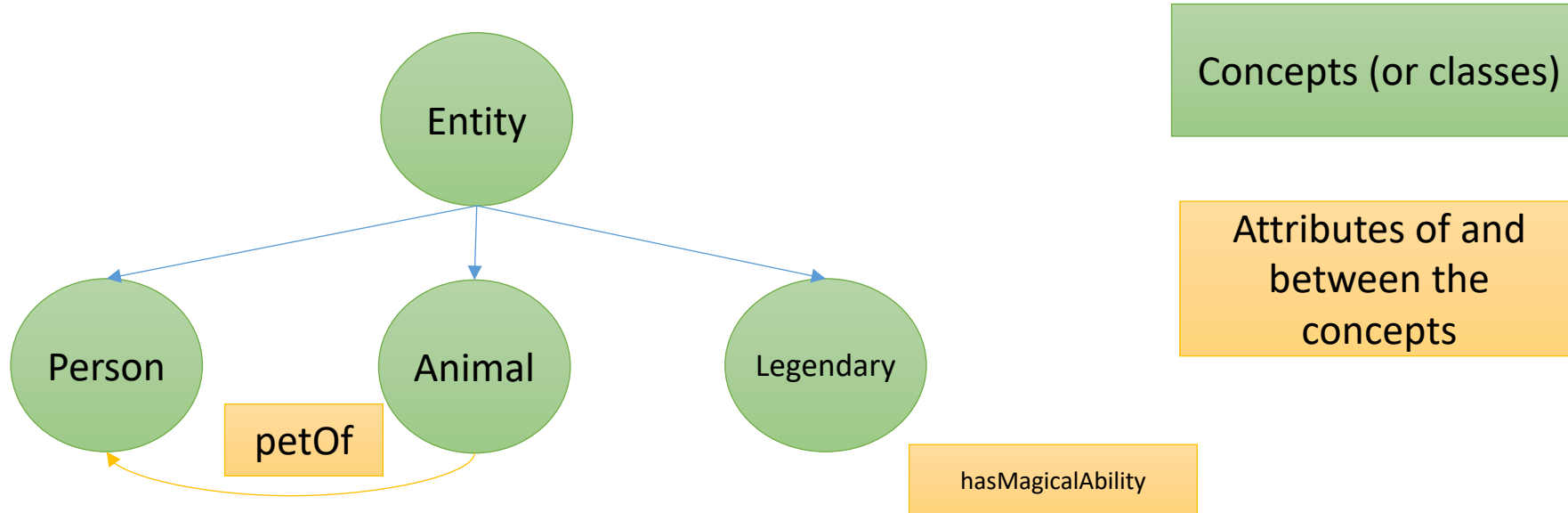
- A small example for characters



Concepts (or classes)

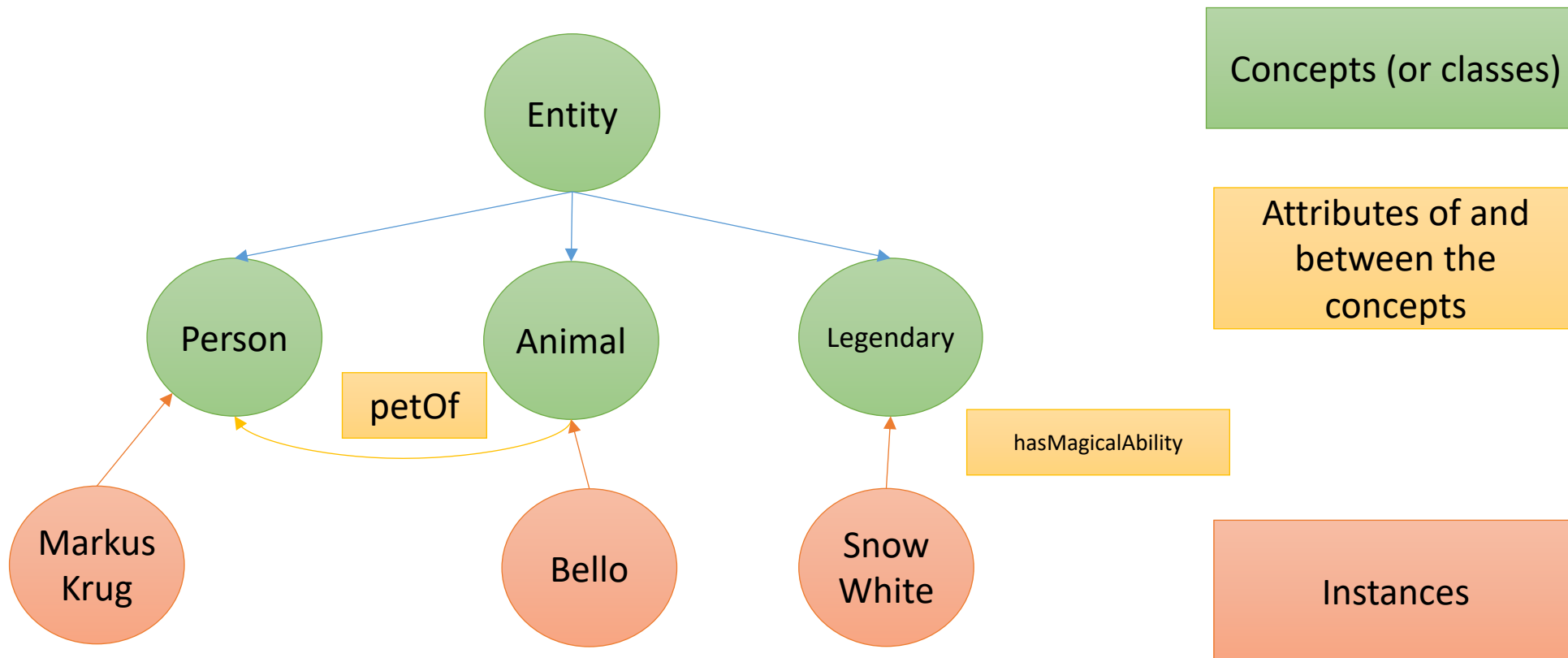
Information Extraction - Ontology

- A small example:



Information Extraction - Ontology

- A small example:



Information Extraction – Knowledge Base

- A small example:

