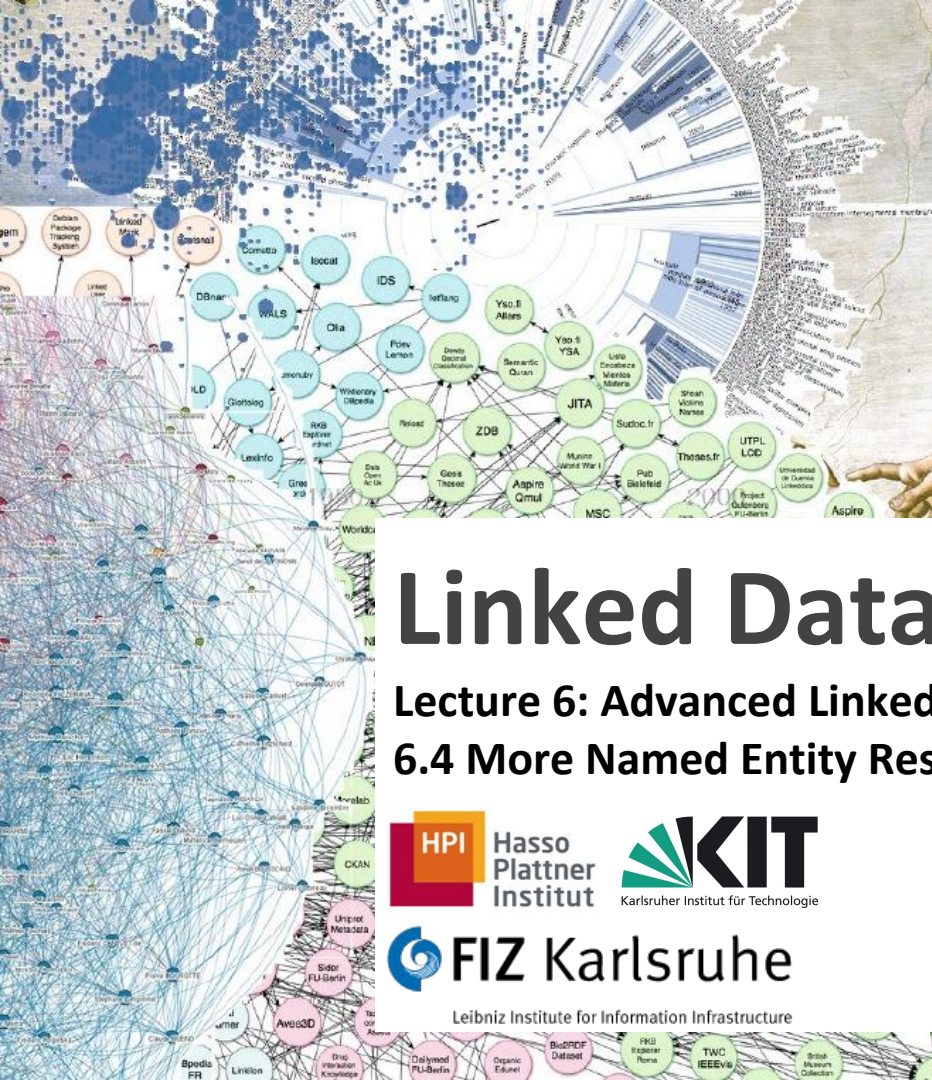




# Linked Data Engineering

## Lecture 6: Advanced Linked Data Based Applications

### 6.4 More Named Entity Resolution



Leibniz Institute for Information Infrastructure

Prof. Dr. Harald Sack  
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure  
Karlsruhe Institute of Technology

**Autumn 2016**

# Named Entity Resolution

**Neil Armstrong**, the 38-year-old civilian commander,  
radios to earth and the mission control room here:  
“Houston, Tranquility Base here, The Eagle has landed.”

[http://dbpedia.org/resource/Neil\\_Armstrong](http://dbpedia.org/resource/Neil_Armstrong)

- 1) Detect Named Entities in Text
- 2) Determine possible Candidate Entities
- 3) Filter Entity Candidates
- 4) Disambiguate Entity Candidates according to Context

# (1) Detect Named Entities in Text

## Named Entity Resolution

- Linguistic Analysis with
  - **POS Tagging**
  - Named Entity Recognition
  - n-gram Analysis
  - Normalization

Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the mission control room here: "Houston, Tranquility Base here, The Eagle has landed."



Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the mission control room here : "Houston, Tranquility Base here, The Eagle has landed."

Adjective

Adverb

Conjunction

Determiner

Noun

Number

Preposition

Pronoun

Verb

<http://parts-of-speech.info/>

# (1) Detect Named Entities in Text

## Named Entity Resolution

- Linguistic Analysis with
  - POS Tagging
  - **Named Entity Recognition**
  - n-gram Analysis
  - Normalization

Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the mission control room here: “Houston, Tranquility Base here, The Eagle has landed.”

Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the mission control room here: “Houston, Tranquility Base here, The Eagle has landed.”

-  Person
-  Location
-  Organization

<http://nlp.stanford.edu:8080/ner/>

# (1) Detect Named Entities in Text

## Named Entity Resolution

- Linguistic Analysis with
  - POS Tagging
  - Named Entity Recognition
  - **n-gram Analysis**
  - Normalization

Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the **mission control room** here: “Houston, Tranquility Base here, The Eagle has landed.”



|         |         |
|---------|---------|
| mission | 1-grams |
| control |         |
| room    |         |

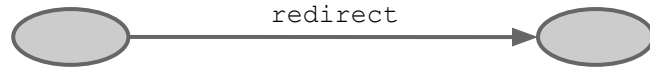
|                 |         |
|-----------------|---------|
| mission control | 2-grams |
| control room    |         |

|                      |         |
|----------------------|---------|
| mission control room | 3-grams |
|----------------------|---------|

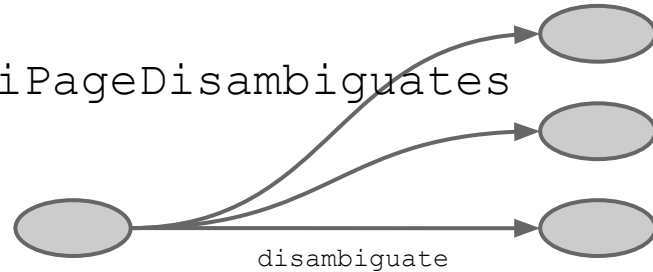
# (2) Determine Possible Entity Candidates

## Named Entity Resolution

- Gazetteer: **semantic entities** ↔ **assigned labels**
- ca. **40M labels** for 9M DBpedia entities
- **Synonyms**
  - redirected via `dbo:wikiPageRedirects`



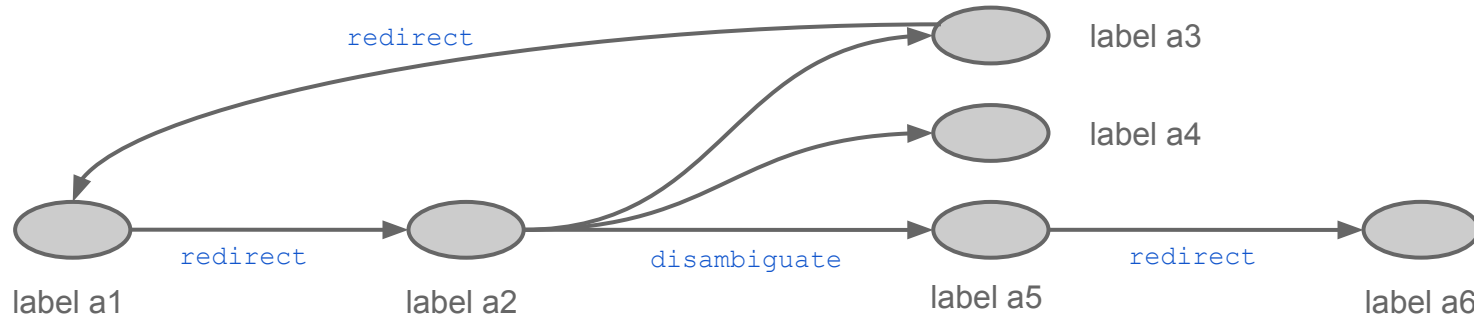
- **Homonyms**
  - disambiguation pages via `dbo:wikiPageDisambiguates`



## (2) Determine Possible Entity Candidates

### Named Entity Resolution

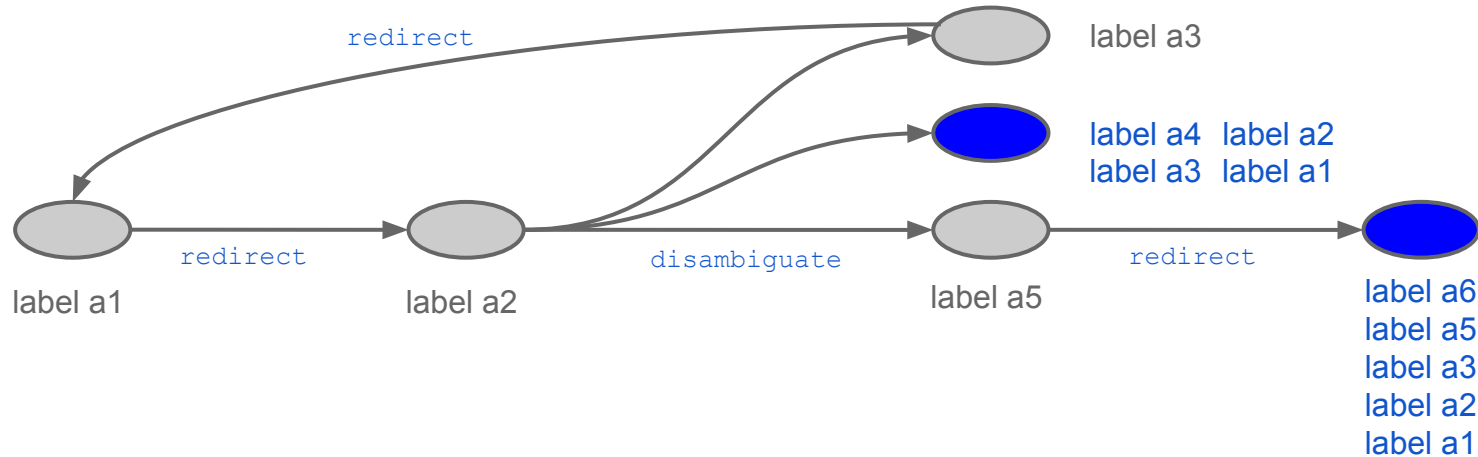
- resolve **chains** and **cycles** in DBpedia link graph
- aggregate all labels from redirect and disambiguation paths within the leafs



## (2) Determine Possible Entity Candidates

### Named Entity Resolution

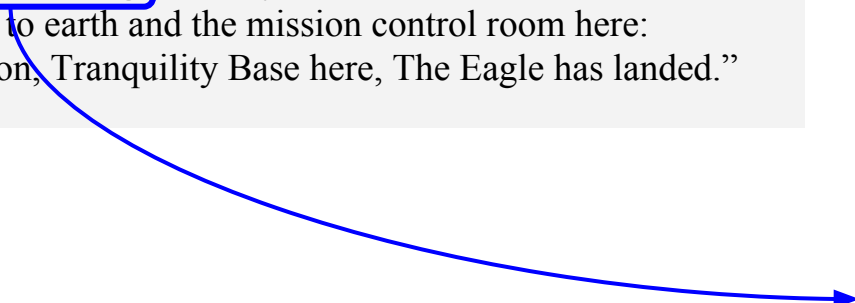
- resolve **chains** and **cycles** in DBpedia link graph
- aggregate all labels from redirect and disambiguation paths within the leafs



# (2) Determine Possible Entity Candidates

## Named Entity Resolution

**Neil Armstrong**, the 38-year-old civilian commander,  
radioes to earth and the mission control room here:  
“Houston, Tranquility Base here, The Eagle has landed.”



| CandidateResources                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="http://wikidata.dbpedia.org/resource/Q1615">http://wikidata.dbpedia.org/resource/Q1615</a>                                             |
| <a href="http://dbpedia.org/resource/Neil_Armstrong">http://dbpedia.org/resource/Neil_Armstrong</a>                                             |
| <a href="http://wikidata.dbpedia.org/resource/Q594654">http://wikidata.dbpedia.org/resource/Q594654</a>                                         |
| <a href="http://wikidata.dbpedia.org/resource/Q6988246">http://wikidata.dbpedia.org/resource/Q6988246</a>                                       |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey)">http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey)</a>                   |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(Ice_hockey_official)">http://dbpedia.org/resource/Neil_Armstrong_(Ice_hockey_official)</a> |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey_official)">http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey_official)</a> |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Air_and_Space_Museum">http://dbpedia.org/resource/Neil_Armstrong_Air_and_Space_Museum</a>   |
| <a href="http://wikidata.dbpedia.org/resource/Q689196">http://wikidata.dbpedia.org/resource/Q689196</a>                                         |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Airport">http://dbpedia.org/resource/Neil_Armstrong_Airport</a>                             |
| <a href="http://wikidata.dbpedia.org/resource/Q3015988">http://wikidata.dbpedia.org/resource/Q3015988</a>                                       |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Achievement">http://dbpedia.org/resource/Neil_Armstrong_Achievement</a>                     |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Hall_of_Engineering">http://dbpedia.org/resource/Neil_Armstrong_Hall_of_Engineering</a>     |
| <a href="http://dbpedia.org/resource/Neil_armstrong_air_and_space_museum">http://dbpedia.org/resource/Neil_armstrong_air_and_space_museum</a>   |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Elementary_School">http://dbpedia.org/resource/Neil_Armstrong_Elementary_School</a>         |
| <a href="http://dbpedia.org/resource/Neil_armstrong">http://dbpedia.org/resource/Neil_armstrong</a>                                             |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Lunar_Outpost">http://dbpedia.org/resource/Neil_Armstrong_Lunar_Outpost</a>                 |

# (3) Filter Entity Candidates

## Named Entity Resolution

Neil Armstrong, the 38-year-old civilian commander, radioes to earth and the mission control room here: “Houston, Tranquility Base here, The Eagle has landed.”

- Person
- Location
- Organization

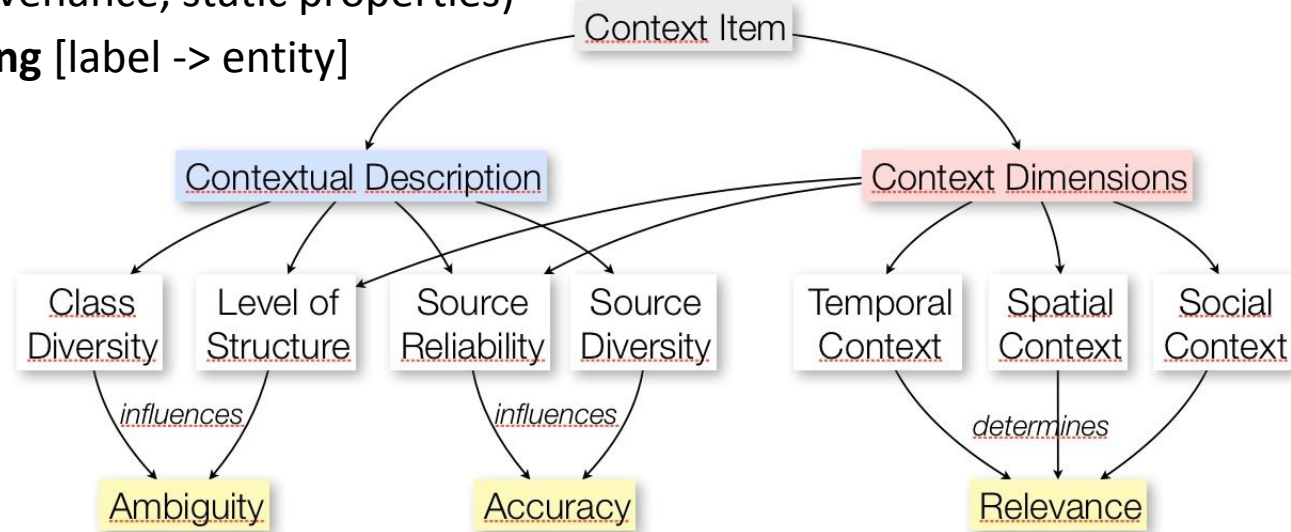
| CandidateResources                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="http://wikidata.dbpedia.org/resource/Q1615">http://wikidata.dbpedia.org/resource/Q1615</a>                                             |
| <a href="http://dbpedia.org/resource/Neil_Armstrong">http://dbpedia.org/resource/Neil_Armstrong</a>                                             |
| <a href="http://wikidata.dbpedia.org/resource/Q594654">http://wikidata.dbpedia.org/resource/Q594654</a>                                         |
| <a href="http://wikidata.dbpedia.org/resource/Q6988246">http://wikidata.dbpedia.org/resource/Q6988246</a>                                       |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey)">http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey)</a>                   |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(Ice_hockey_official)">http://dbpedia.org/resource/Neil_Armstrong_(Ice_hockey_official)</a> |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey_official)">http://dbpedia.org/resource/Neil_Armstrong_(ice_hockey_official)</a> |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Air_and_Space_Museum">http://dbpedia.org/resource/Neil_Armstrong_Air_and_Space_Museum</a>   |
| <a href="http://wikidata.dbpedia.org/resource/Q689196">http://wikidata.dbpedia.org/resource/Q689196</a>                                         |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Airport">http://dbpedia.org/resource/Neil_Armstrong_Airport</a>                             |
| <a href="http://wikidata.dbpedia.org/resource/Q3015988">http://wikidata.dbpedia.org/resource/Q3015988</a>                                       |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Achievement">http://dbpedia.org/resource/Neil_Armstrong_Achievement</a>                     |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Hall_of_Engineering">http://dbpedia.org/resource/Neil_Armstrong_Hall_of_Engineering</a>     |
| <a href="http://dbpedia.org/resource/Neil_armstrong_air_and_space_museum">http://dbpedia.org/resource/Neil_armstrong_air_and_space_museum</a>   |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Elementary_School">http://dbpedia.org/resource/Neil_Armstrong_Elementary_School</a>         |
| <a href="http://dbpedia.org/resource/Neil_armstrong">http://dbpedia.org/resource/Neil_armstrong</a>                                             |
| <a href="http://dbpedia.org/resource/Neil_Armstrong_Lunar_Outpost">http://dbpedia.org/resource/Neil_Armstrong_Lunar_Outpost</a>                 |

# (4) Disambiguate Entity Candidates

## Named Entity Resolution

- **Context Analysis**

- considers **ambiguity, accuracy, reliability**
- of **source data** (provenance, static properties)
- as well as of **mapping** [label -> entity]



N.Steinmetz, H.Sack: Semantic Multimedia Information Retrieval Based on Contextual Descriptions, ESWC 2013

# (4) Disambiguate Entity Candidates

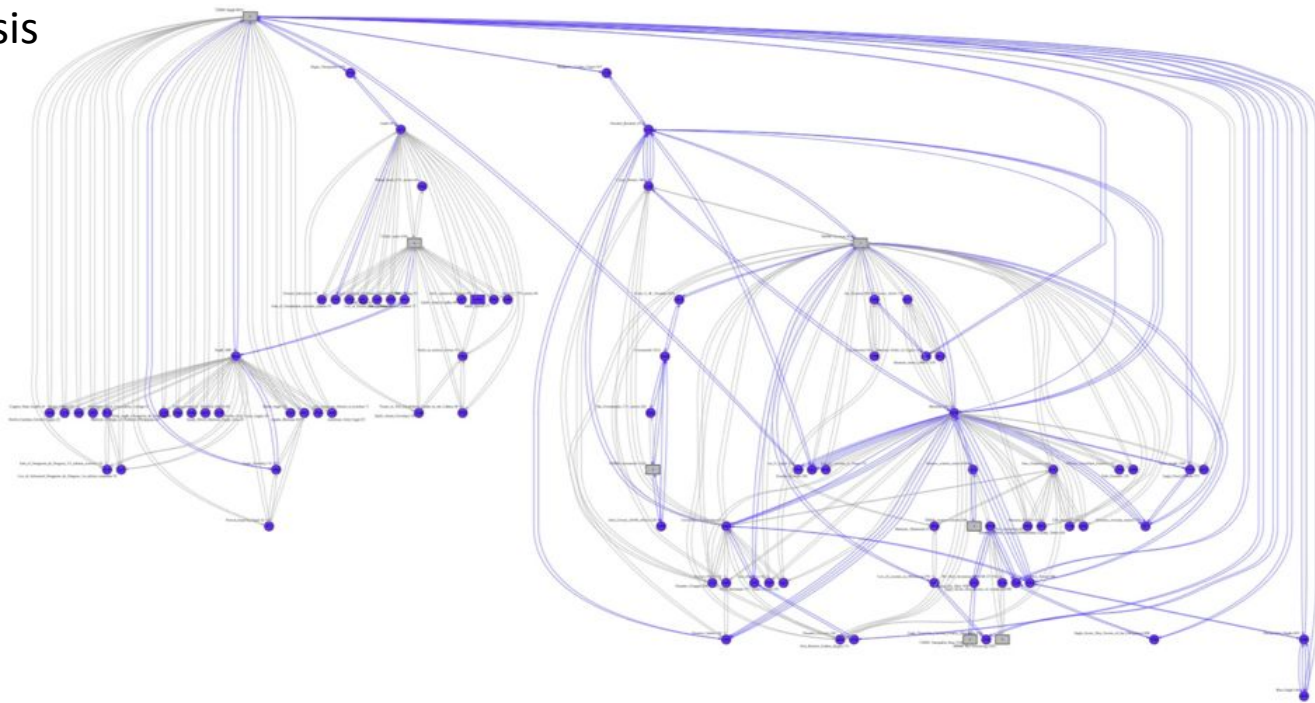
## Named Entity Resolution

- Link Graph Analysis
- Co-Occurrence Analysis
- Relevance Ranking

Neil Armstrong, the 38-year-old civilian **commander**,  
radioes to **earth** and the **mission control room** here:  
“**Houston, Tranquility Base** here, the **Eagle** has landed.”



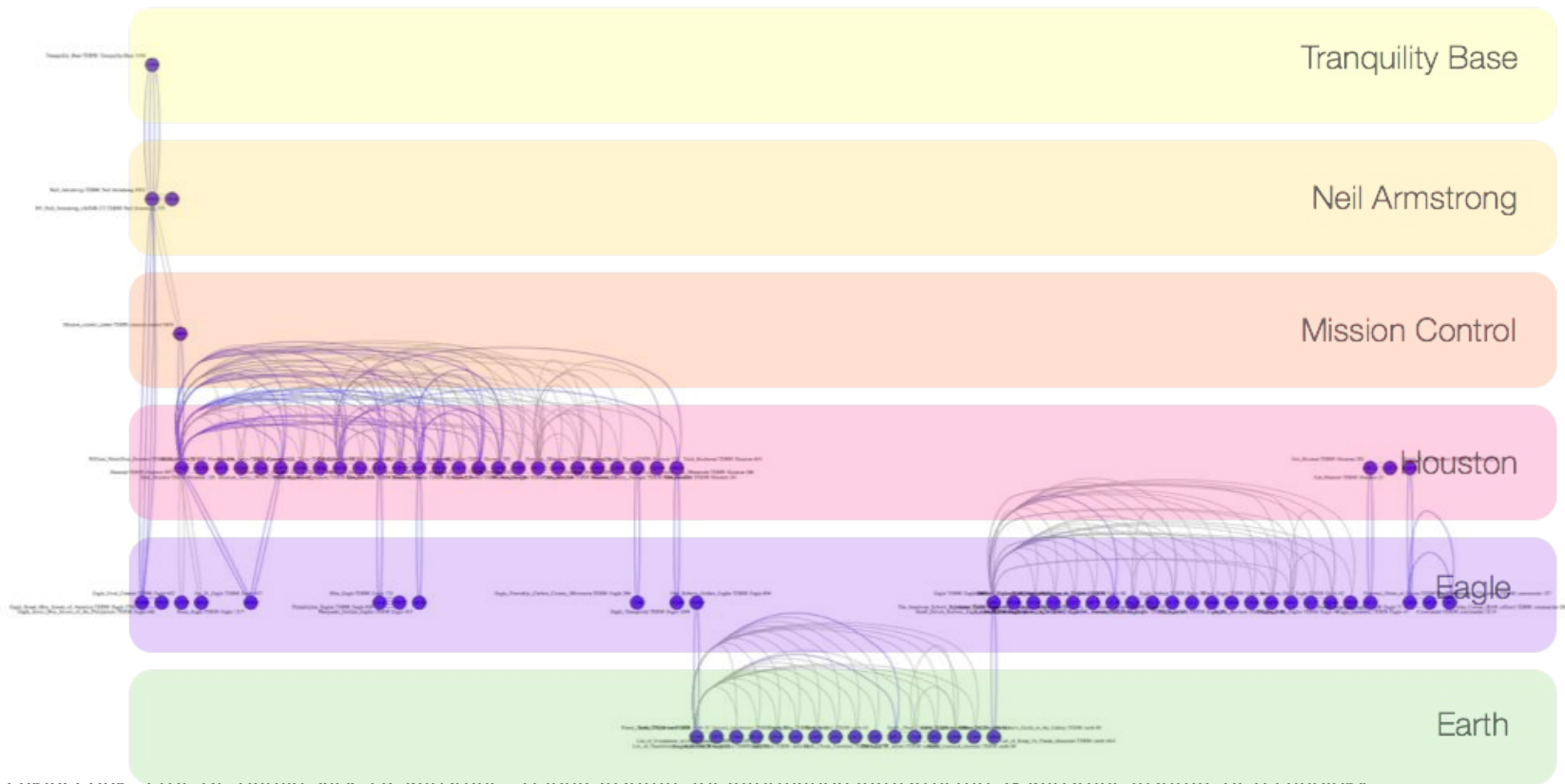
induced Link Graph



# (4) Disambiguate Entity Candi

## Named Entity Resolution

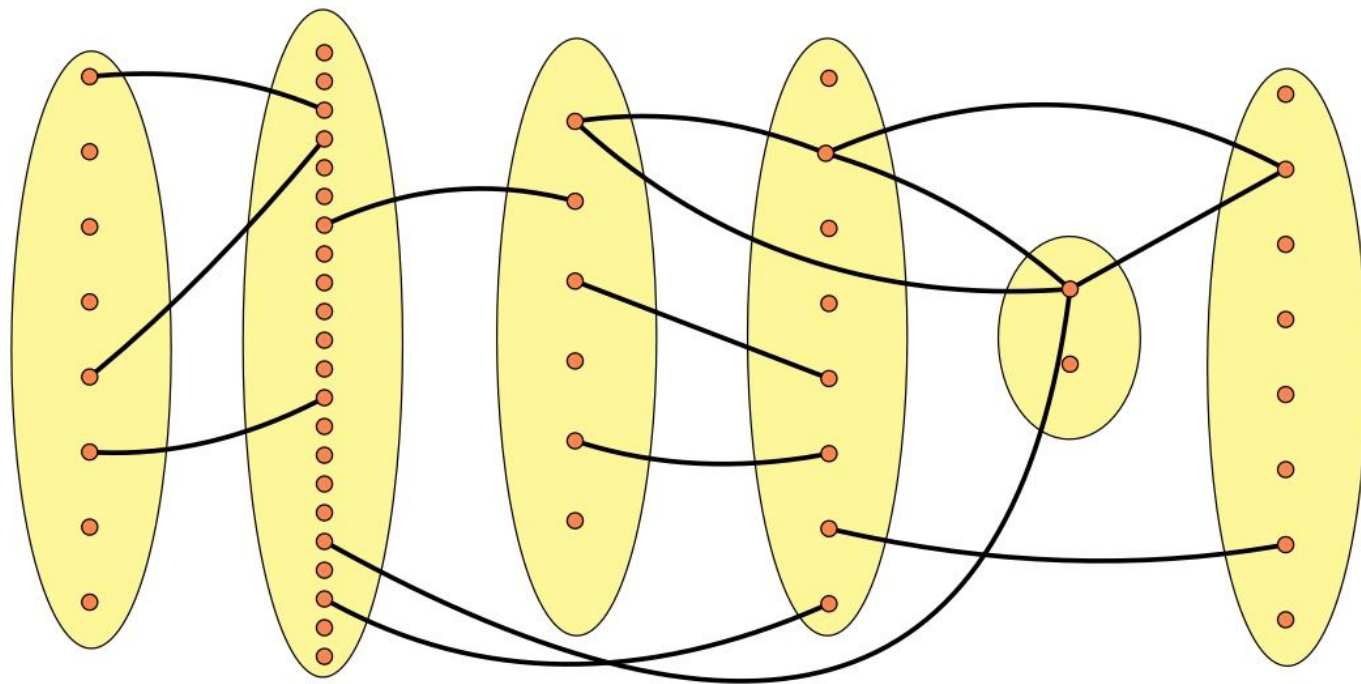
Neil Armstrong, the 38-year-old civilian **commander**,  
radioes to **earth** and the **mission control room** here:  
“**Houston**, **Tranquility Base** here, the **Eagle** has landed.”



# (4) Disambiguate Entity Candidates

Named Entity Resolution

Eagle earth Neil Armstrong Houston Tranquility Base mission control

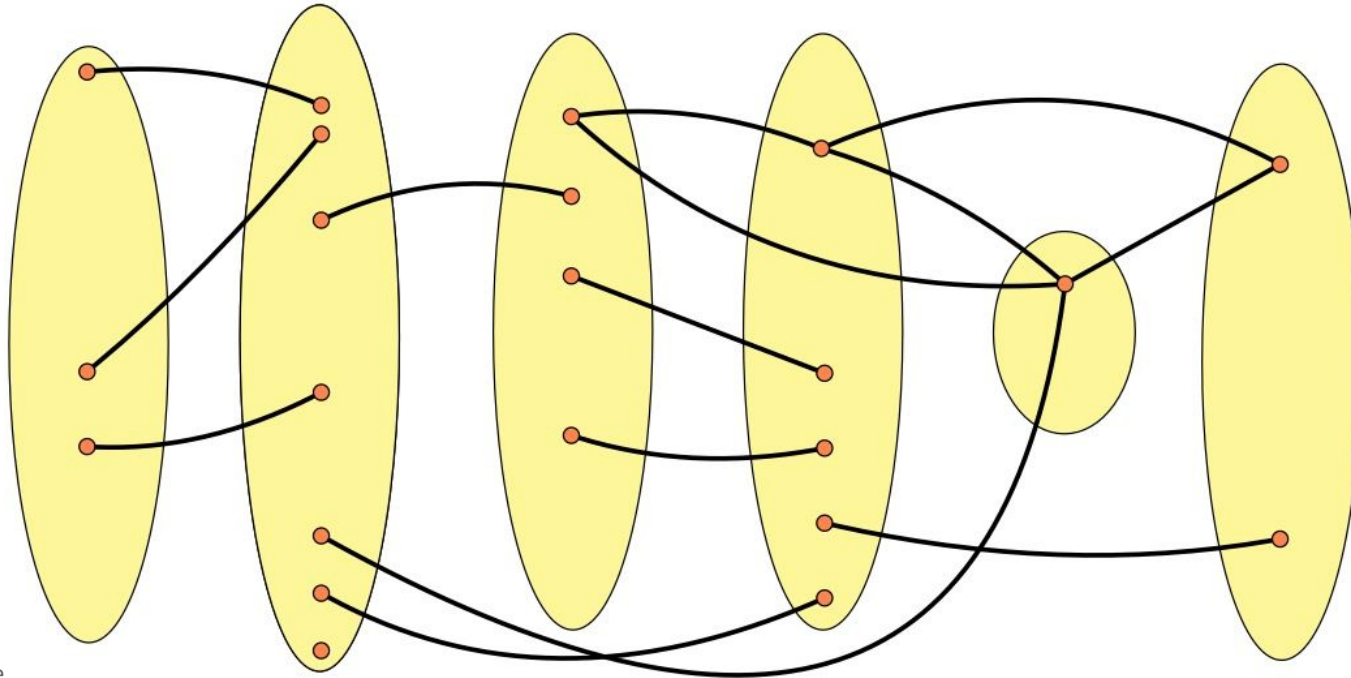


Connected  
Component  
Analysis

# (4) Disambiguate Entity Candidates

Named Entity Resolution

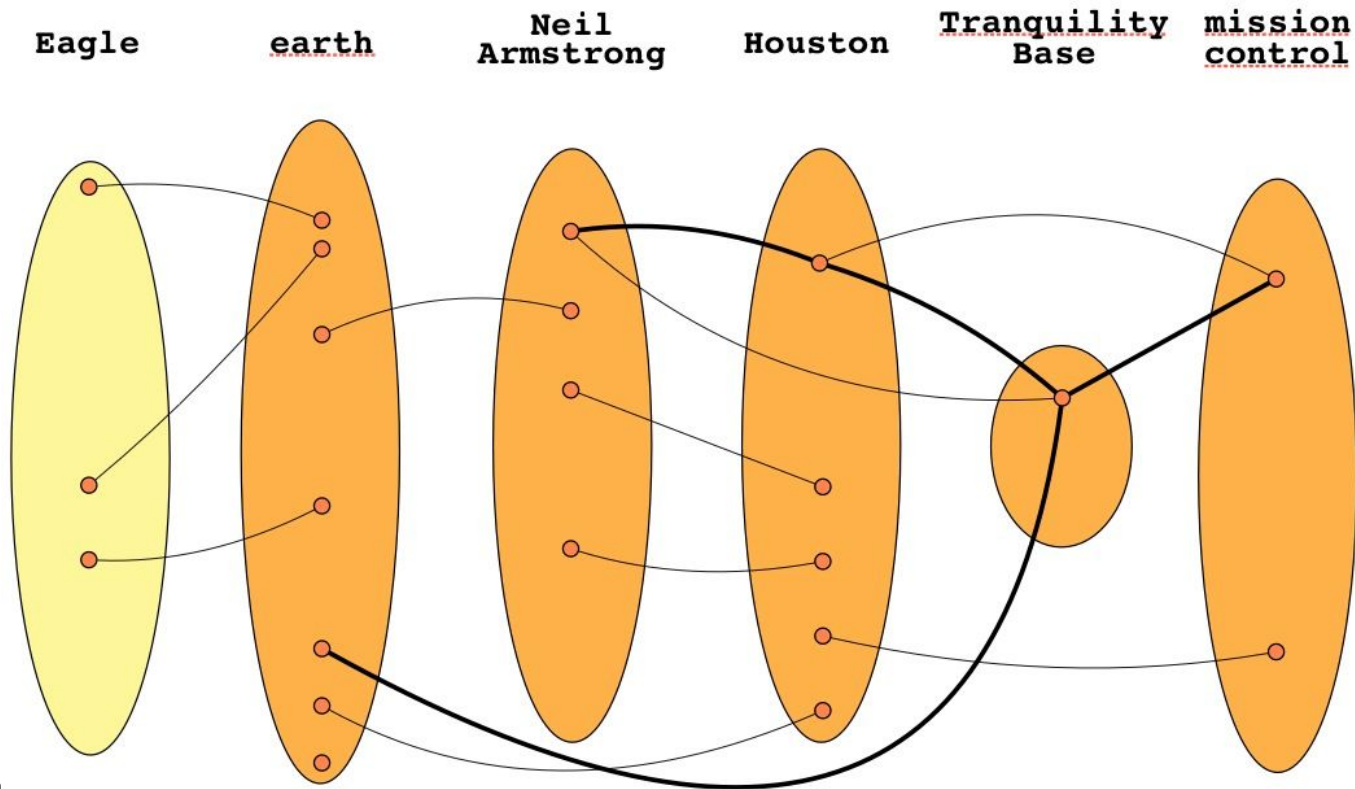
Eagle earth Neil Armstrong Houston Tranquility Base mission control



1) Identify  
Connected  
Components

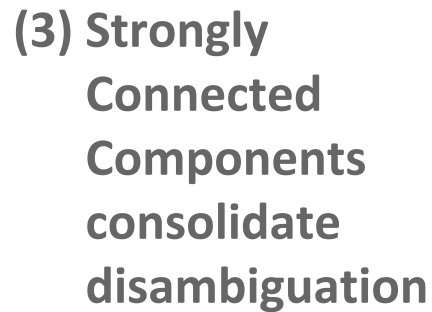
# (4) Disambiguate Entity Candidates

Named Entity Resolution



(2) Identify  
Connected  
Components  
that cover  
most term  
partitions

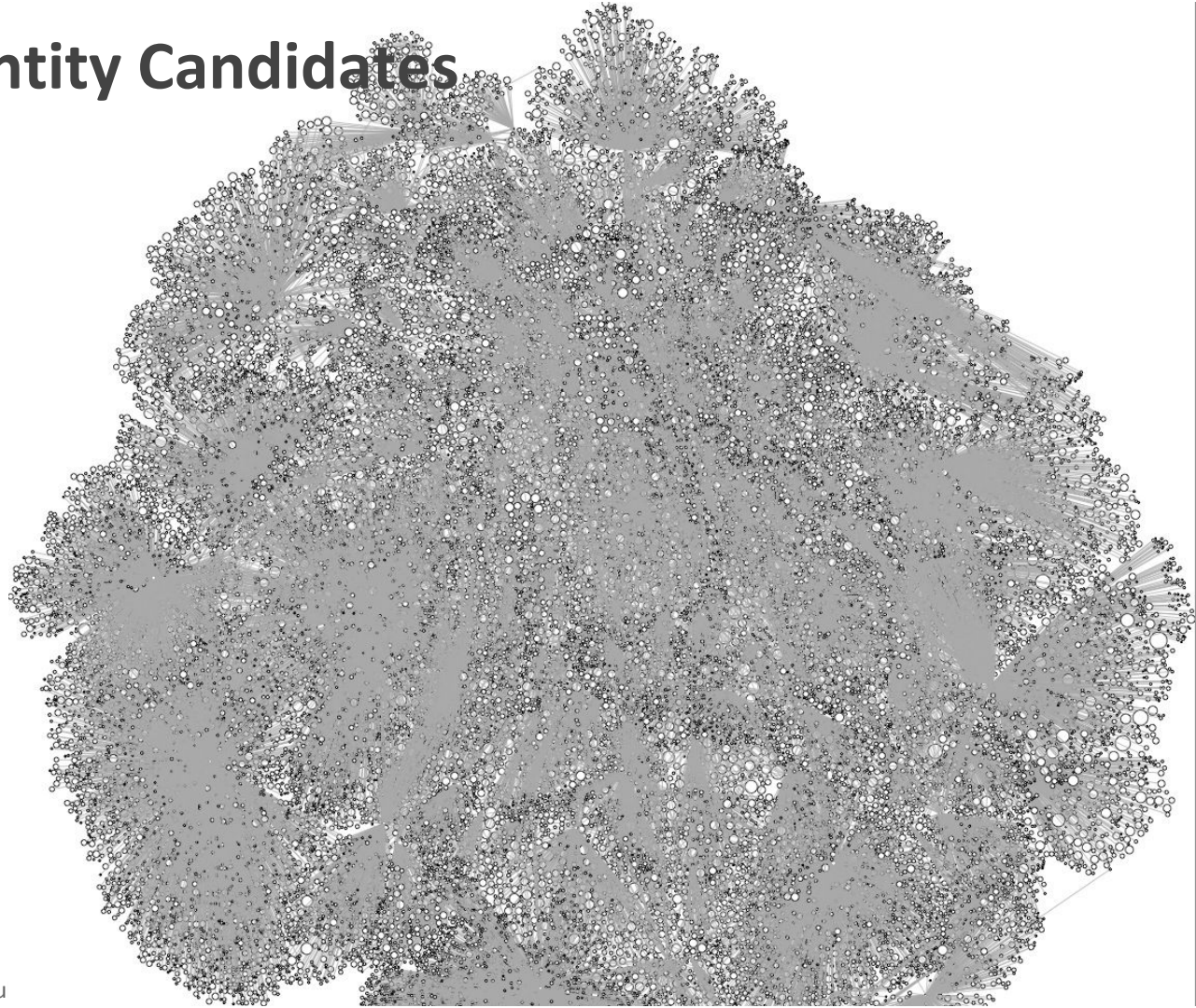
## Named Entity Resolution



# (4) Disambiguate Entity Candidates

Named Entity Resolution

- **Link Graph Analysis**
- Co-Occurrence Analysis
- Relevance Ranking



# (4) Disambiguate Entity Candidates

## Named Entity Resolution

- Link Graph Analysis
- Co-Occurrence Analysis
- Relevance Ranking

Neil Armstrong, the 38-year-old civilian **commander**, radioes to **earth** and the **mission control room** here:  
“Houston, **Tranquility Base** here, The **Eagle** has landed.”

### Neil Armstrong

From Wikipedia, the free encyclopedia

For other people named Neil Armstrong, see Neil Armstrong (disambiguation).

**Neil Alden Armstrong** (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics High-Speed Flight Station, now known as the Dryden Flight Research Center, where he logged over 900 flights. He later completed graduate studies at the University of Southern California.

A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight, as command pilot of Gemini 8, in 1966, becoming NASA's first civilian astronaut to fly in space. On this mission, he performed the first docking of two spacecraft, with pilot David Scott.<sup>[1]</sup>

Armstrong's second and last spaceflight was as mission commander of the Apollo 11 Moon landing, in July 1969. On this mission, Armstrong and Buzz Aldrin descended to the lunar surface and spent two and a half hours exploring, while Michael Collins remained in lunar orbit in the Command Module. Along with Collins and Aldrin, Armstrong was awarded the Presidential Medal of Freedom by President Richard Nixon; President Jimmy Carter presented Armstrong the Congressional Space Medal of Honor in 1978; he and his former crewmates received the Congressional Gold Medal in 2009.

Armstrong died in Cincinnati, Ohio, on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery.<sup>[2][3]</sup>



### Neil Armstrong (ice hockey)

From Wikipedia, the free encyclopedia

For other people named Neil Armstrong, see Neil Armstrong (disambiguation).

**Neil Armstrong** (born December 20, 1932) is a former professional ice hockey linesman, and is an Honoured Member of the Hockey Hall of Fame.

#### Career <sup>[edit]</sup>

Armstrong began playing minor hockey in Galt, Ontario, but he never did go beyond that. He was offered a chance to officiate a game in the same league. Armstrong accepted and later earned his Ontario Hockey Association certification.

He officiated his first National Hockey League game on November 17, 1957, when he was 24. In the game, which was between the Boston Bruins and Toronto Maple Leafs, the two teams got into a brawl near the end of the game. Armstrong broke up a fight involving Fern Flaman, who later skated up to him with his arm dangling and proclaimed "you broke my arm!". However it turned out that Flaman was only kidding.

During his career, he had only been seriously injured once and had never missed any games, which helped him gain the nickname "ironman". His one major injury came in 1971 when Philadelphia Flyers player Gary Dornhoefer fell along the boards, and knocked Armstrong up against the glass. Dornhoefer's stick cut Armstrong's hand and broke a bone, forcing him to wear a cast for three months. On October 16, 1973, Armstrong was honoured in a ceremony at the Detroit Olympia for officiating his 1,314th game, which broke the previous record set by George Hayes.<sup>[1]</sup>

In total, Armstrong officiated a total of 1,744 games and retired in 1978.<sup>[2]</sup> After retiring, he became a scout for the Montreal Canadiens. He was inducted into the Hockey Hall of Fame as an official in 1991.

His son, Doug Armstrong, later became General Manager of the Dallas Stars. Following his tenure with Dallas, Doug became the Executive VP and General Manager of the St. Louis Blues and continues to serve in that capacity.<sup>[3]</sup>

| Neil Armstrong    |                                                                   |
|-------------------|-------------------------------------------------------------------|
| <b>Born</b>       | Neil Armstrong<br>December 20, 1932 (age 82)<br>Plympton, ON, CAN |
| <b>Occupation</b> | Former NHL linesman                                               |

# (4) Disambiguate Entity Candidates

## Named Entity Resolution

### General Approach:

1. Always start disambiguation with the most reliable algorithm on most accurate and most reliable data
2. Resolve the remaining ambiguity with less reliable algorithms on less reliable data

### KEA Named Entity Resolution

1. Connected Component Analysis on Link Graph
2. Co-occurrence on wikipedia text corpus
3. Popularity Based Link Graph Analysis
4. Negative Context Analysis

*N. Steinmetz, H. Sack: Semantic Multimedia Information Retrieval Based on Contextual Descriptions, ESWC 2013*  
*N. Steinmetz, H. Sack: About the Influence of Negative Context. ICSC 2013*  
*R. Usbeck et al, GERBIL - General Entity Annotator Benchmark, WWW 2014*  
*J. Waitelonis and H. Sack, Named Entity Linking in #Tweets with KEA, #Microposts 2016*



**Next: 05 Semantic Search**

**OpenHPI - Course Linked Data Engineering - Lecture 6: Advanced Linked Data Based Applications**