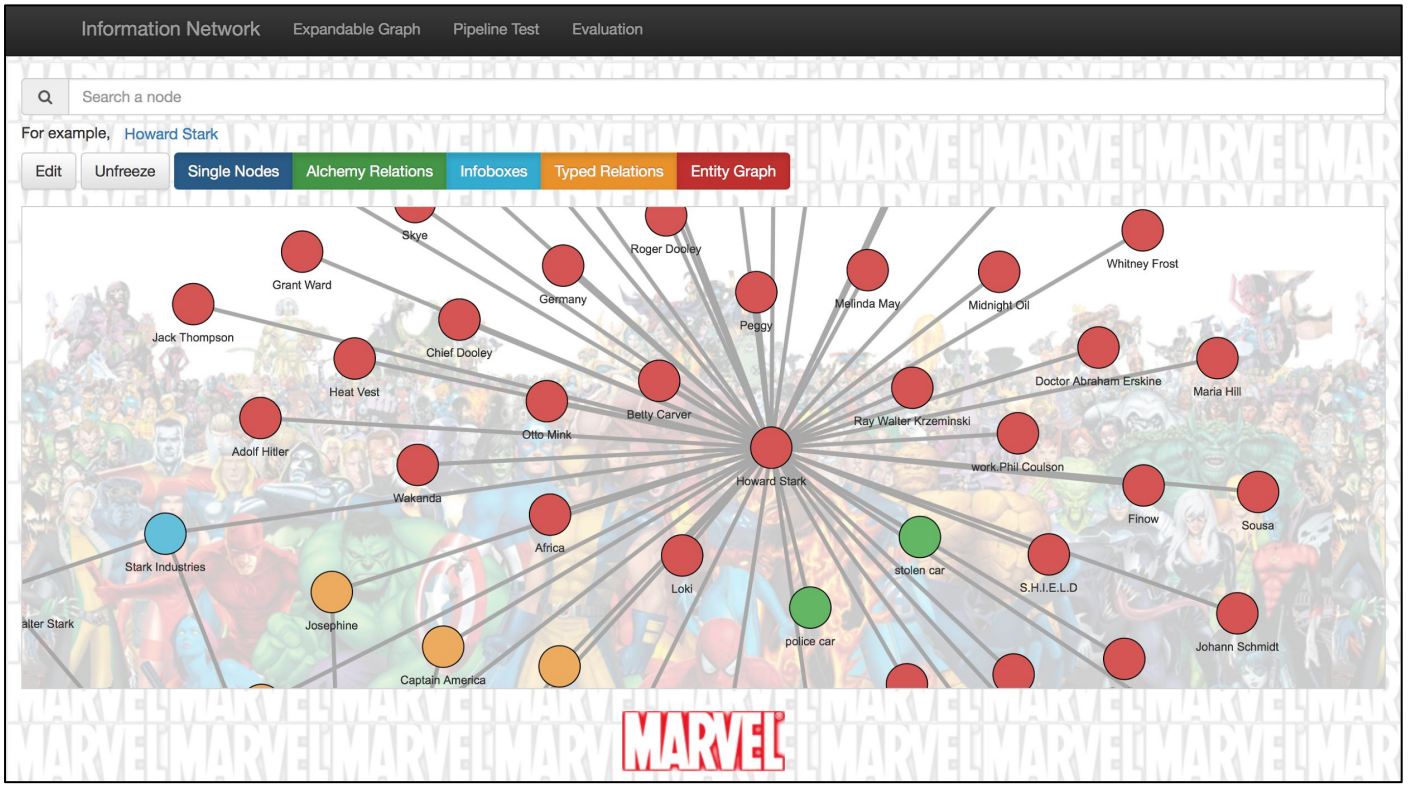
Question Answering Technologies behind and with IBM Watson

TECHNISCHE UNIVERSITÄT DARMSTADT

Information Network - Final Presentation



Problem Statement - Introduction

content is often strongly dependent among each other and is spread over many years

- ***In which relation stand this two persons ?***
- ***Where plays the Book/Film/Series ?***
- ***In which Book/Film/Series did this person occurs ?***
- ***Is the Person already dead ?***
- ***When was the first time the Person/Group occurs ?***
- ***Is the person a member of a group ?***
- ***What persons/cities/organisations/country's occurs?***
- ***What are the different roles of an actor ?***

Problem Statement - Introduction

content is often strongly dependent among each other and is spread over many years

- ***In which relation stand this two persons ?***
- ***Is the Person dead/alive ?***
- ***Is the person a member of a group ?***
- ***What is the meaning of an citie/organisation/country's and where it occure?***
- ***What are the different roles of an actor***

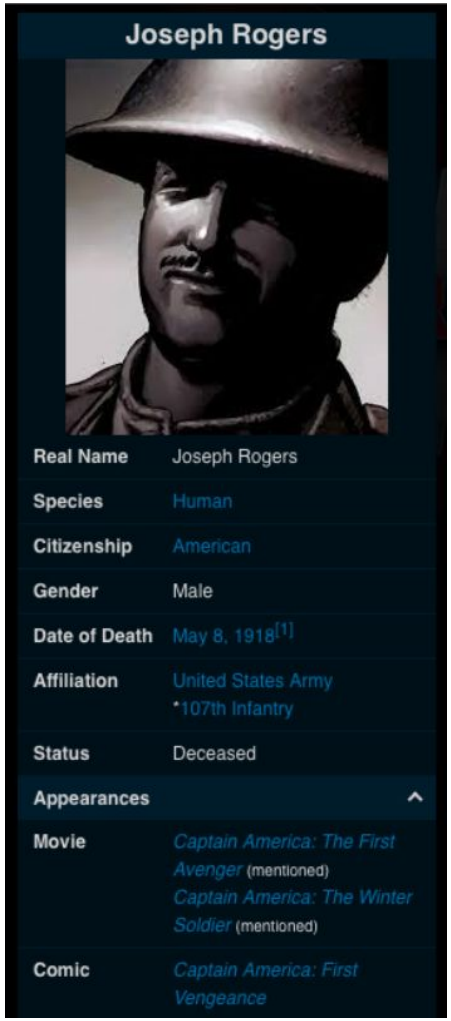
1) Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- 7367 crawled html files

- extract informations from the html files -> 13.3MB Json files containing text
  - Text: Extract all … and combine them to one text
  - Infoboxes
  - Title
  - URL

2) Infobox relations




```
<h2>Infobox Joseph Rogers:</h2>
    <table>
        <tr><td>Real Name</td><td>Joseph Rogers</td></tr>
        <tr><td>Species</td><td>Human</td></tr>
        <tr><td>Citizenship</td><td>American</td></tr>
        <tr><td>Gender</td><td>Male</td></tr>
        <tr><td>Date of Death</td><td>May 8, 1918[1]</td></tr>
        <tr><td>Affiliation</td><td>United States Army, *107th Infantry</td></tr>
        <tr><td>Status</td><td>Deceased</td></tr>
        <tr><td>Movie</td><td>Captain America: The First Avenger (mentioned), Captain Am
        <tr><td>Comic</td><td>Captain America: First Vengeance</td></tr>
    </table>
</section>
```

```
[
    {"edge":"Real Name","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Joseph Rogers"},
    {"edge":"Species","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Human"},
    {"edge":"Citizenship","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"American"},
    {"edge":"Gender","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Male"},
    {"edge":"Date of Death","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"May 8, 1918"},
    {"edge":"Affiliation","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"United States Army"},
    {"edge":"Affiliation","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"107th Infantry"},
    {"edge":"Status","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Deceased"},
    {"edge":"Movie","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Captain America: The First Avenger (
    {"edge":"Movie","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Captain America: The Winter Soldier
    {"edge":"Comic","nodeOne":"Joseph Rogers","type":"infobox","nodeTwo":"Captain America: First Vengeance"}
]
```

2) Alchemy Language Service

TECHNISCHE UNIVERSITÄT DARMSTADT

- Services for the analysis and interpretation of the content and context of texts on web pages, in news articles etc.
- Uses a predefined domain model
  - This could be individually learnt with the Watson Knowledge Studio
  - We used a standard model for english text content

Note:

- Services allow different input sizes
- these sizes may vary even after the input type
- Limited service calls
- short texts are a problem -> can't detect the language and throws an error

2.a) Alchemy Entity Extraction -> Nodes

TECHNISCHE UNIVERSITÄT DARMSTADT

- Extract all entities of a text
  - 42 primary types: *City, Organization, Person*
  - 976 subtypes: C.-Airport,C.-Building,
    O.Newspaper, O-Politician,
    P.-Actor, P.-Chef, P.-Celebrity
- Count: how often he noticed an entity e.g.
  - Joseph Rogers -> count = 1
  - Detect that e.g “joseph” or “he” in the context
    stand for Joseph Rogers -> count = 2
- Relevance:
  - depicts the significance of each unique term
  - The higher the relevance score, the more important that term to the central meaning of the document
- Some failures:
  - Mixing different entities, e.g. Chan.Â Raina
  - Extract the wrong type, see Purpel Heart Medal


```
{
  "count":7,
  "text":"Joseph Rogers",
  "type":"Person",
  "relevance":0.904197
},
{
  "count":4,
  "text":"Steve Rogers",
  "type":"Person",
  "relevance":0.581804
},
{
  "count":3,
  "text":"Sarah Rogers",
  "type":"Person",
  "relevance":0.369227
},
{
  "count":1,
  "text":"107th Infantry Regiment",
  "type":"Organization",
  "relevance":0.243262
},
{
  "count":2,
  "text":"World War",
  "type":"FieldTerminology",
  "relevance":0.213352
},
{
  "count":1,
  "text":"United States Armed Forces",
  "type":"Organization",
  "relevance":0.211032
},
{
  "count":1,
  "text":"United States Army",
  "type":"Organization",
  "relevance":0.176365
},
{
  "count":1,
  "text":"Purple Heart Medal",
  "type":"Organization",
  "relevance":0.154761
},
```


2.b) Alchemy Relations Extraction

TECHNISCHE UNIVERSITÄT DARMSTADT

Extract SOA Relations: Subject - Object - Action

```
{
    "sentence":" After Joseph died, Sarah had to raise Steve by herself, lamenting that
                 Joseph left them too soon, and noting that Steve was very much like his
                 father, wanting to follow Joseph's steps as a soldier.",
    "subject":{"text":"Sarah"},
    "action":{"lemmatized":"have to raise","verb":{"text":"raise","tense":"past"},
              "text":"had to raise"},
    "object":{"text":"Steve"}
}
```

Generate
Relation for Graph

Simple string matching for the subject and object with the alchemy entities that were found in the same file

```
{
    "subject":  "Sarah Rogers",
    "action":   "had to raise",
    "object":   "Steve Rogers"
}
```

Note:

- you can restrict the service, that the response contain only elements that contains at least one entity. This has removed a lot of the garbage in response.

2.c) Alchemy Typed Relations

TECHNISCHE UNIVERSITÄT DARMSTADT


```
{
  "sentence":"Joseph fought with the 107th during World War I dying as a consequence
              of mustard gas attack before his son was born.",
  "score":"0.824973",
  "arguments":[
    {
      "entities":[
        {
          "id":"-E0",
          "text":"Joseph Rogers",
          "type":"Person"
        }
      ],
      "part":"first",
      "text":"his"
    },
    {
      "entities":[
        {
          "id":"-E1",
          "text":"Steve Rogers",
          "type":"Person"
        }
      ],
      "part":"second",
      "text":"son"
    }
  ],
  "type":"parentOf"
}
```


- identify different types of connections between detected entities
- try to recognize the entities
- the types are created with the creation of a model
  - Watson Knowledge Studio

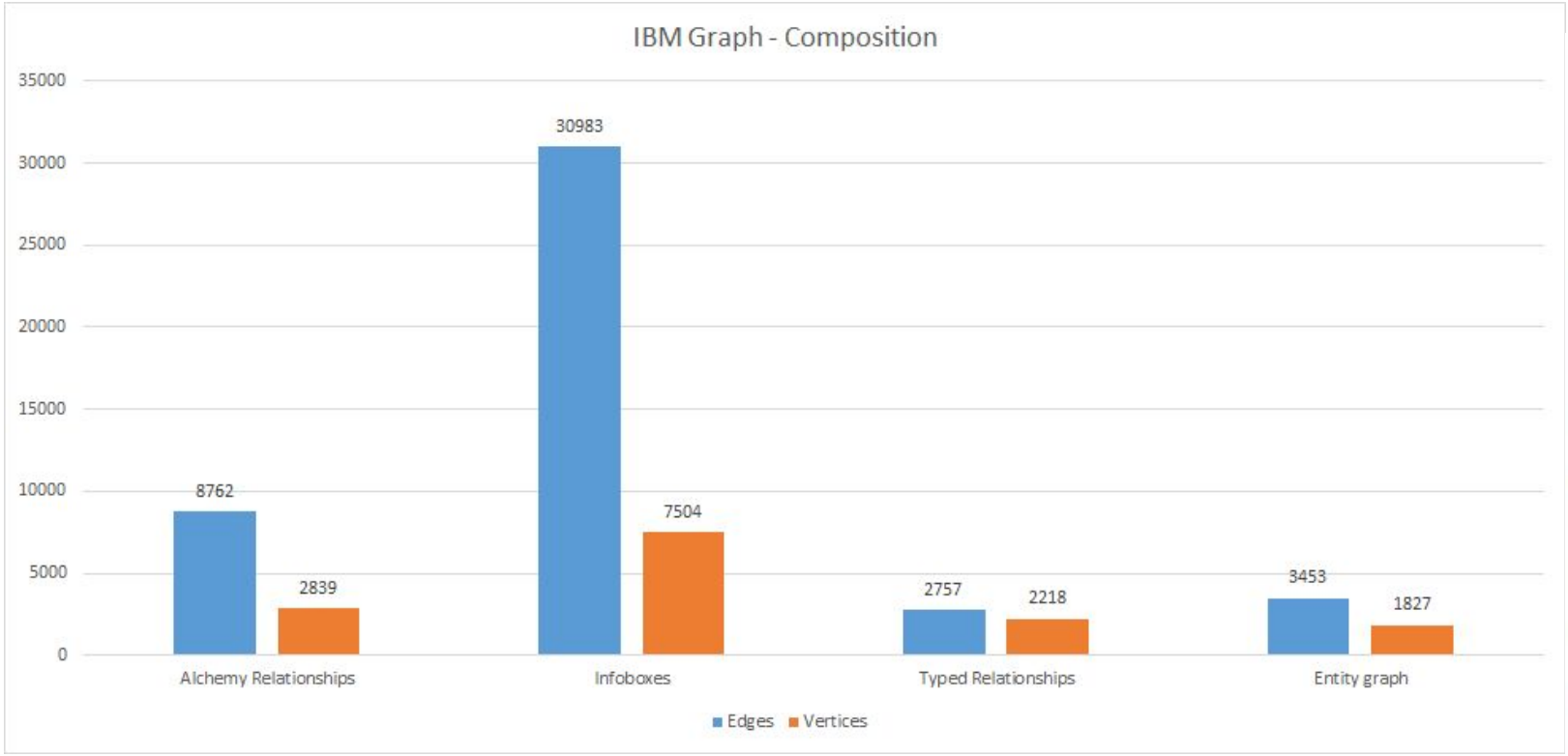
3) Graph DB

TECHNISCHE UNIVERSITÄT DARMSTADT

- We use IBM Graph as a Database
  - Based on Apache TinkerPop
  - Used to store nodes and edges
  - Uses Gremlin Query Language
- What we store:

Nodes	Edges
Label	Label
Relevance-score from Alchemy	Alchemy relationship score
Vertex count	Edge count
List of URLs and Titles	Relations mode
	List of Sentences, URLs and Titles

4) IBM Graph - Composition



5) Web UI

TECHNISCHE UNIVERSITÄT DARMSTADT

- Show the different Graphs we extract from the services
  - Entity graph: entities that are connected together occur in the same file
  - Tagged Relation Graph: used the relation that are defined in the Watson Knowledge Studio model
  - Relation Graph: used the SOA-Relations
- Edit function -> create a groundtruth
  - Nodes & Edges can be:
    - Added
    - Removed
    - Changed

6) Demo

7) Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

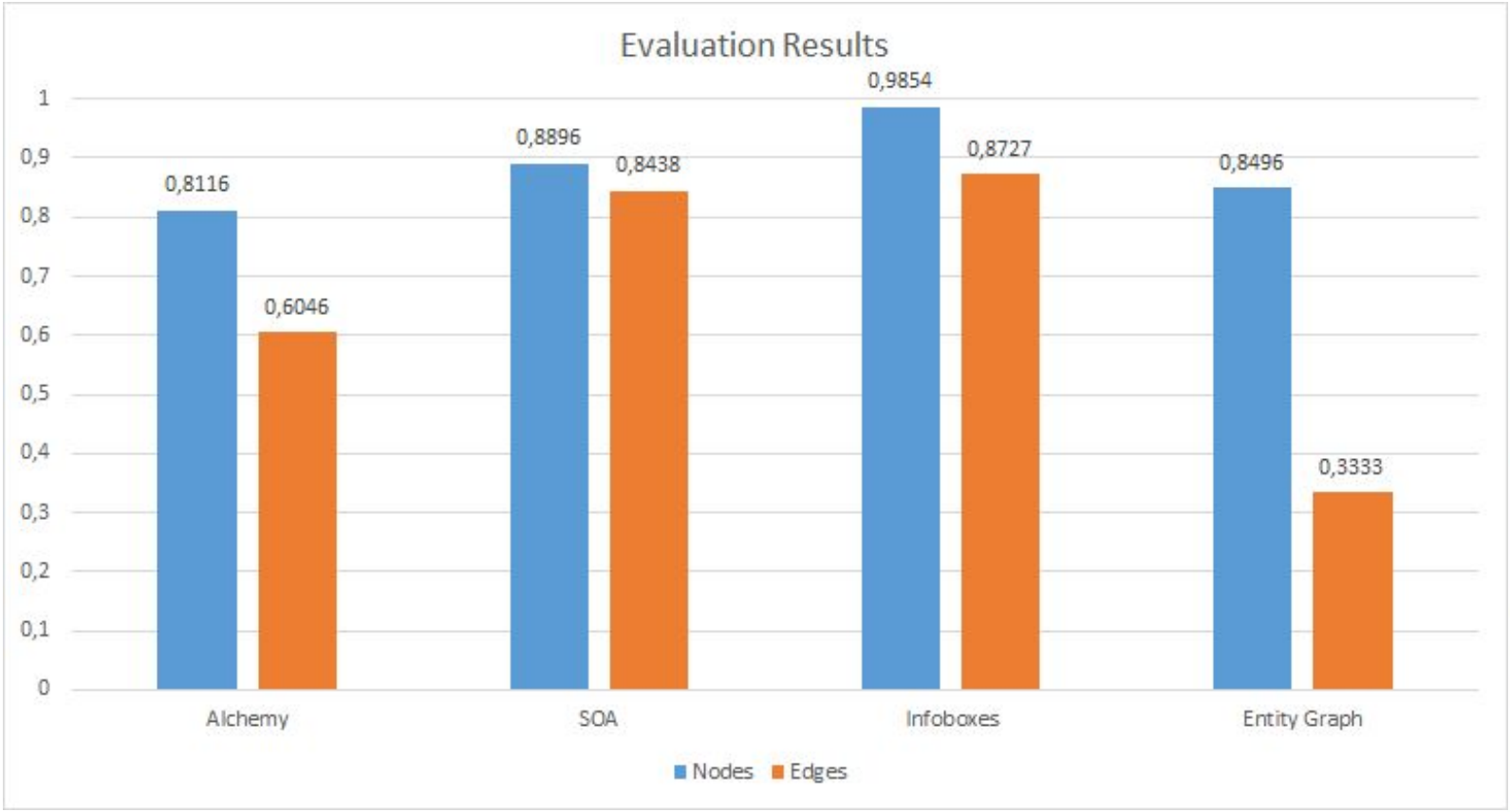
Qualitative Evaluation

150 Edges and their 300 corresponding Nodes

Picked randomly

Repeated 4 times for each Service.

7) Evaluation Feature Results



8) Future Work

TECHNISCHE UNIVERSITÄT DARMSTADT

Seasonal Graphs - so you don't spoiler

Entity Linking

Better String Matching approach

Filtering: Typ's, Scores etc.

Thank you!