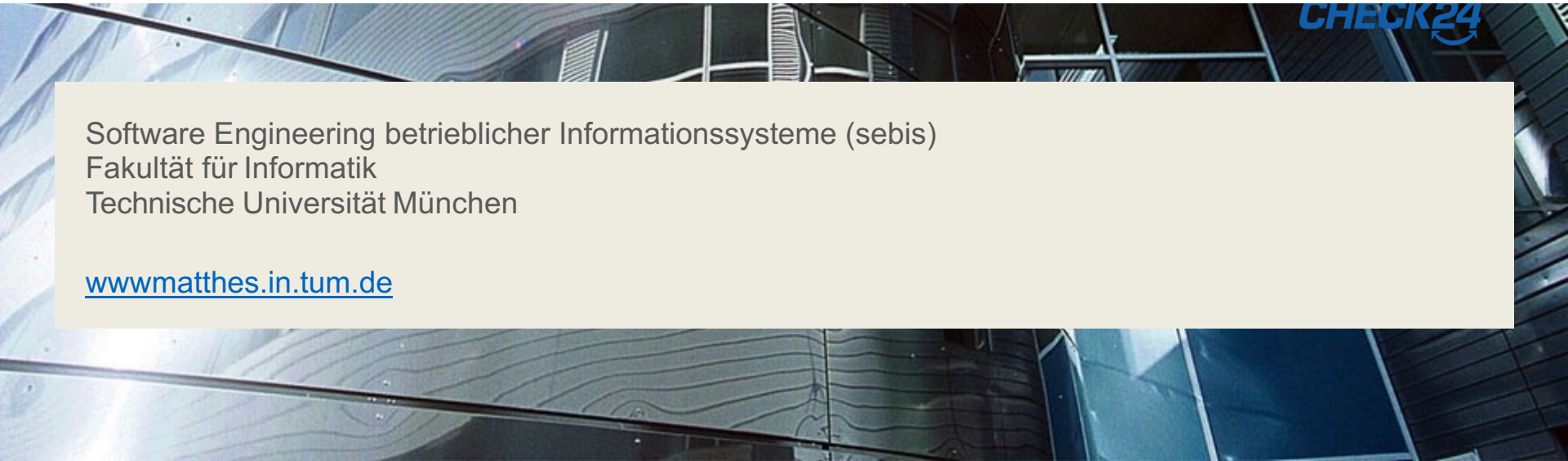
# Using Web Annotations to Represent Relations between Structured and Unstructured Information in Semantic Wikis

**Master thesis – Final presentation**

Shivguru Rao BhimasenaRao VisweswaraRao

*Advisor:* Daniel Braun

Munich, 09.07.2018




Software Engineering betrieblicher Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München

wwwmatthes.in.tum.de

# Agenda

1. Problem Statement – A Recap
2. Solution - Web-Annotations
3. Solution – JSON-LD Standard Extension
4. Demo
5. System Architecture
6. Revisiting Research questions
7. Conclusion

# Problem Statement – A recap

## Unstructured Data

### Master's Thesis Bhimasenarao Shivguru Rao

Last modified Dec 21, 2017

No tags assigned

### Using Web Annotations to Represent Relations between Structured and Unstructured Information in Semantic Wikis

Hybri Wikis combine unstructured information in the form of Wiki texts with structured information. However, there is no link between the structured and the unstructured information. Therefore, if a piece of information is represented as both, structured and unstructured information, it has to be update manually in both formats in order to keep the content consistent.

Aim of this thesis is to use the metamodel of SocioCortex and Machine Learning to automatically link structured and unstructured data in Hybrid Wikis and also extract structured information from unstructured content. A web application should be built in order to allow users to manually annotate unstructured data from SocioCortex in order to create training data for the Machine Learning algorithm and visualize the links between the data format.

## Structured Data

### Attributes of this Student Project

| | |
|---|---|
| Title (de) | Nutzung von Web Annotations zur Repräsentation von Relationen zwischen strukturierten und unstrukturierten Informationen in Semantic Wikis |
| Title (en) | Using Web Annotations to Represent Relations between Structured and Unstructured Information in Semantic Wikis |
| Project | |
| Type | Master's Thesis |
| Status | started |
| Student | |
| Advisor | Daniel Braun |
| Supervisor | Prof. Dr. Florian Matthes |

# Problem Statement – A recap

## Unstructured Data

### Master's Thesis Bhimasenarao Shivguru Rao

Last modified Dec 21, 2017

No tags assigned

**Using Web Annotations to Represent Relations between Structured and Unstructured Information in Semantic Wikis**

Hybri Wikis combine unstructured information in the form of Wiki texts with structured information. However, there is no link between the structured and the unstructured information. Therefore, if a piece of information is represented as both, structured and unstructured information, it has to be update manually in both formats in order to keep the content consistent.

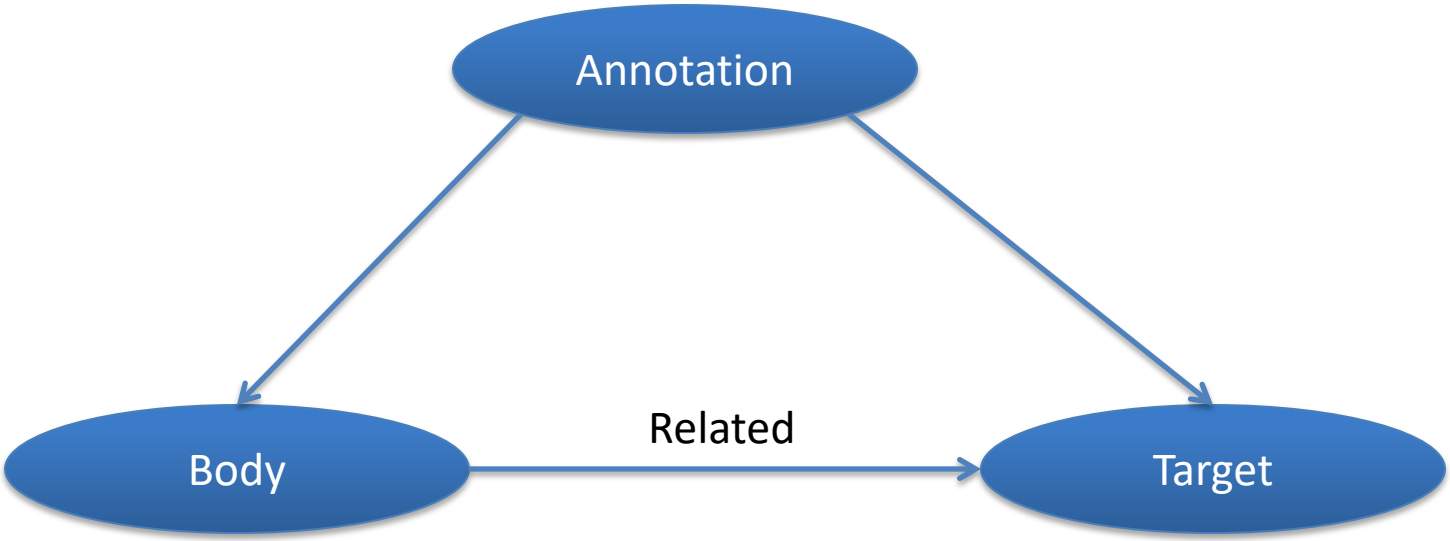
Aim of this thesis is to use the metamodel of SocioCortex and Machine Learning to automatically link structured and unstructured data in Hybrid Wikis and also extract structured information from unstructured content. A web application should be built in order to allow users to manually annotate unstructured data from SocioCortex in order to create training data for the Machine Learning algorithm and visualize the links between the data format.

## Structured Data

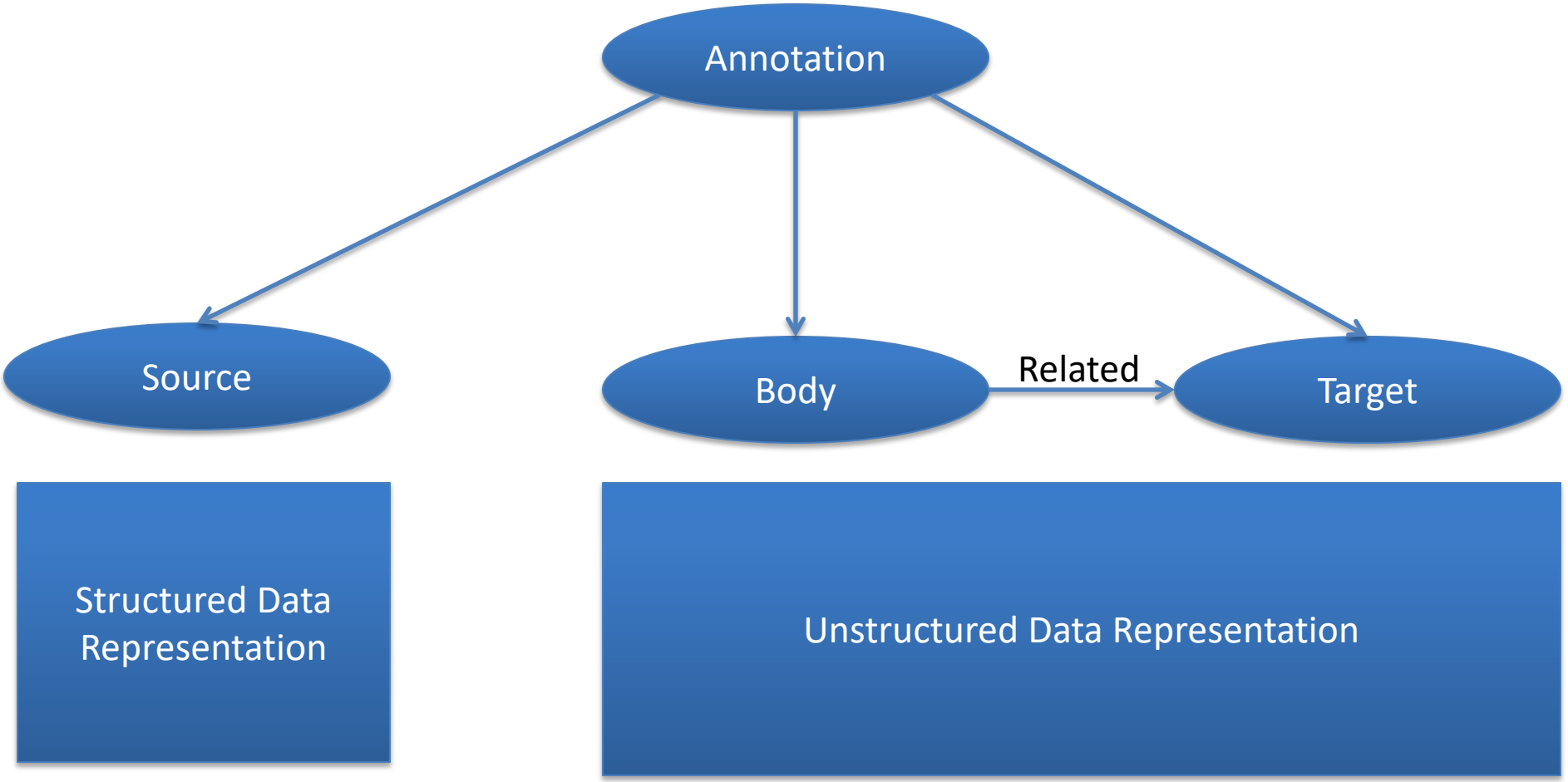
### Attributes of this Student Project

| | |
|---|---|
| Title (de) | Nutzung von Web Annotations zur Repräsentation von Relationen zwischen strukturierten und unstrukturierten Informationen in Semantic Wikis |
| Title (en) | Using Web Annotations to Represent Relations between Structured and Unstructured Information in Semantic Wikis |
| Project | |
| Type | Master's Thesis |
| Status | started |
| Student | |
| Advisor | Daniel Braun |
| Supervisor | Prof. Dr. Florian Matthes |

# Solution -Web Annotation



# Solution -Web Annotation for linking

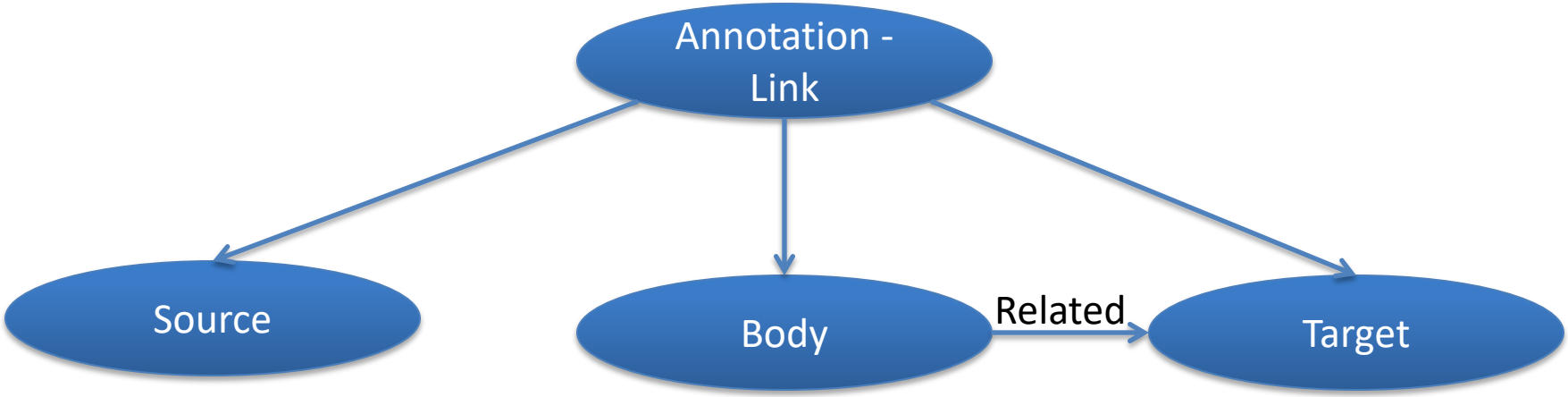


# Solution – JSON LD Standard

- Method on encoding linked data using JSON
- Designed Around concept of a “context” to provide additional mappings
- W3C provides “anno.jsonld” - Meta data for creating web annotations

```
{
    "@context": {
        "name": "http://xmlns.com/foaf/0.1/name",
            "homepage": {
                "@id": "http://xmlns.com/foaf/0.1/workplaceHomepage",
                "@type": "@id"
            },
        "Person": "http://xmlns.com/foaf/0.1/Person"
    },
    "@id": “qsm234234nsdw23”,
    "@type": "Person",
    "name": "John Smith",
    "homepage": "https://www.example.com/"
}
```

# Solution - Standard Extension




```
"source": {
    "id": "1t4b4rniixb8s",
    "value": 2017,
    "parent": {
        "id": "1rn347knpl6ii",
        "name": "Q2-2017"
    },
    "DataPositionSelector": 0,
    "type": "Inter Entity"
}
```

DOM Manipulation:

```
<span id="link_2" >
    2017
</span>
```

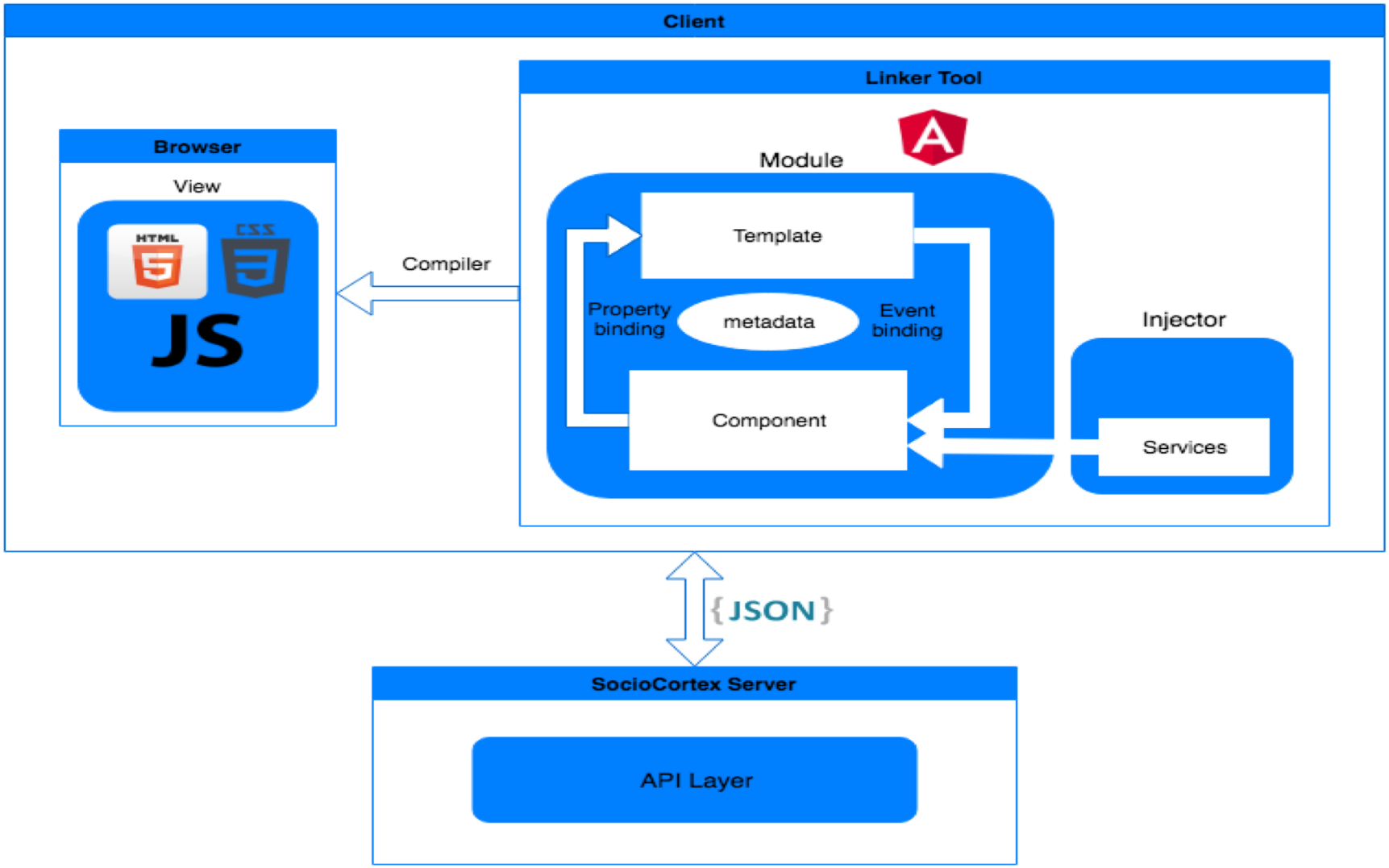
JSON Content:

```
"body": {
    "value": "2017 ",
    "format": "text/plain",
    "type": "TextualBody"
}
```

```
"target": {
    "id": "c7gt17zhps0l",
    "selector": {
        "value": "span#link_2",
        "type": "CssSelector"
    }
}
```

# Demo

# Architecture



# Revisiting Research questions

- What should be the format of the link?
- Are there any annotation standards to link data?
- How can we store and retrieve links?
- How to detect and handle the changes?

# Conclusion and Future Work

- Handling link creation for data spanning across multiple HTML elements
- Handling Link changes from with in the tool
- Automatic Link Creation – A Recommender system approach

# Thank You