Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

Kenan Mahic

Impact of clarifying uncertainty in graph-based entity disambiguation

Masteroppgave i Datateknologi
Veileder: Krisztian Balog
August 2020

NTNU
Kunnskap for en bedre verden

Kenan Mahic

Impact of clarifying uncertainty in graph-based entity disambiguation

Masteroppgave i Datateknologi
Veileder: Krisztian Balog
August 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

NTNU
Kunnskap for en bedre verden

This thesis is dedicated to my loving family. My mother Edhija Mahic, my brother Safurudin Mahic, my sister-in-law Sanda Mahic, my niece Una Mahic, Kerim Canovic and my friends. Thank you for all the love, support and guidance throughout my life and education. I would also like to extend a special thank you to my academic supervisor Krisztian Balog at the Department of Computer Science at NTNU and UiS for his helpful guidance. Your dynamism and motivation has been an inspiration.

Summary

The goal of this thesis is to survey the impact clarification questions have on graph-based based entity disambiguation. If these clarification questions have an improvement on the performance measures an implementation of these questions might be purposeful for some types of named entity linking systems.

This done through the creation of knowledge graphs commonly used in graph-based entity disambiguation and simulating the effect clarification questions would have on the performance. This was assessed using measures outlined in chapter 4.

The results seem encouraging, in most simulations the clarification question seemed to help the system evaluate to the correct named entity. There are however some concerns. The dataset used might not have been the best to gauge how this would generalize. Further work needs to be done on a more varied dataset to draw absolute conclusions.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Named Entity Linking is not a trivial task due to the name variation and ambiguity problem. Name variation means an entity can be mentioned in different ways. For example, the entity Michael Jeffrey Jordan can be referred to using numerous names, such as Michael Jordan, MJ, and Jordan. Whereas the ambiguity problem is related to the fact that a name may refer to different entities depending on the context. Here is an example(Hoffart et al. (2011)) for the ambiguity problem, the name Bulls can apply to more than one entity in Wikipedia, such as the NBA team Chicago Bulls, the football team Belfast Bulls, etc. In general, a typical entity linking system consists of several modules, namely Mention Detection, Candidate Entity Generation, Candidate Entity Ranking and Disambiguation. A brief description of each module is given below.

1. Mention Detection — The goal is to detect all “linkable” phrases in the document. This step is highly recall oriented. A common approach is building a dictionary of entity surface forms and checking all document n-grams against the dictionary.

2. Candidate Entity Generation — In this module, the NEL system aims to retrieve a set of candidate entities(including name variants) by filtering out the irrelevant entities in the knowledge base. The retrieved set contains all possible entities that may refer to an entity mention.

3. Candidate Entity Ranking — Here, different kinds of evidence are leveraged to rank the candidate entities to find the most likely entity for the mention.

4. Disambiguation/Unlinkable Mention Prediction — This module will validate whether the top-ranked entity identified in the previous module is the target entity for the given mention. If not, then it will return NIL for the mention. It can consider additional types of evidence be it prior importance, contextual similarity, coherence, etc.

1.2 Objectives

From the conclusions of the pre-project conducted as a step towards this goal we can say modern day entity linking systems do not perform adequately. The main goal of this Master's Thesis will therefore be to investigate a new approach to handling the entity linking task for a conversational scenario and have it hopefully perform within acceptable measures. To get us to this goal the initial idea was to split this main goal into several smaller ones. First we wanted to do an error analysis of the pre-project to inform us about requirements for the new approach being made. From there we were to create 3-4 baselines that represent different types of entity linking systems and try to adjust them to the conversational scenario. It was also planned to develop a novel method for clarification questions to remove uncertainty for disambiguation. This was all planned to be evaluated using the same movie corpus that was used during the pre-project.

Chapter 2

Related Work

2.1 Preliminary Error Analysis

2.1.1 Methodology

As the basis for this master's thesis is to produce some sort of improvement in entity linking in conversational settings(be it major or novel), an error analysis of the results from the pre-project would provide useful information in how to advance with this task.

Based on the results from the pre-project the tested systems performance was found lacking. Even though the amount of data that was used fall short of the amount needed to draw decisive conclusions, there is enough data to take a deep dive into the common struggles the tested systems had.

One of the biggest hurdles the systems faced was disambiguating to the film series in question. They would instead disambiguate to a character, comic or just one movie from the franchise. As can be seen in the in figure 2.1, the spot Deadpool is wrongly annotated. It should have been "Deadpool_(film)"

Intuitively one would think the added context of running the more verbose modes would

```
'response': []}, {'index': 20, 'utterance': "Not particularly. I don't really like comic
book movies that much. But I really wanted to see Deadpool, so I'd probably be more willing
to watch that than most comic book movies.", 'response': [{'spot': 'comic book', 'start': 38,
'link_probability': 0.3967929780483246, 'rho': 0.3299468159675598, 'end': 48, 'id': 6231,
'title': 'https://en.wikipedia.org/wiki/Comic_book'}, {'spot': 'Deadpool', 'start': 94,
'link_probability': 1, 'rho': 0.6126736402511597, 'end': 102, 'id': 184420, 'title':
'https://en.wikipedia.org/wiki/Deadpool'}]}, {'index': 21, 'utterance': 'oh what makes
```


Figure 2.1: Example of struggling between a character and movie

help in this regard, but it seemed to only lower the score for 2/3 systems. TagMe being the only exception in most cases with a small increase in its precision, this however would come at the expense of its recall. This trade off is promising in a way as it would appear that for TagMe it starts correctly disambiguating more often with these verbose modes. The problem with TagMe seems to lie in that it tries to add entity links to completely irrelevant parts of the text. As shown in figure 2.2 the common word neat is annotated to the wikipedia link for Near-Earth Asteroid Tracking.


```
'end': 6, 'id': 2654186, 'title': 'https://en.wikipedia.org/wiki/Zodiac_(film)'}]}, {'index':
6, 'utterance': 'And I just think serial killers, in general, are interesting, so the movie
was really good. And it was just neat to see that world. Like it went really in-depth. It was
like almost 3 hours long, so you got to really feel like you were a part of that world and
time period. And see what the detectives and the police did and the investigators.',
'response': [{'spot': 'neat', 'start': 108, 'link_probability': 1, 'rho': 0.5, 'end': 112,
'id': 478364, 'title': 'https://en.wikipedia.org/wiki/Near-Earth_Asteroid_Tracking'},
```


Figure 2.2: Example of wrong mention spot

On a first-look basis it appears the more verbose modes add very little actual context for a big portion of the entities in the best case, and in the worst cases it only seems to confuse the entity linking systems. This effect would appear to worsen with more verbose mode we ran with the full context mode having the most drop off in F1 scores. One interesting point is that the previous-and-current mode would also increase the precision for the google cloud entity linking system at a small dip in recall. This might point to the previous-and-current method having the most promise.

Another problem arises on the opposite end. While some times the systems have difficulties with linking to franchises, in other cases they have major difficulties linking to the correct disambiguation of one movie from a franchise. As demonstrated in figure 2.3.


```
'Not that I can think of.', 'response': []}, {'index': 15, 'utterance': 'how about different
type of movies like Once Upon a Deadpool', 'response': [{'spot': 'Deadpool', 'start': 52,
'link_probability': 1, 'rho': 0.5481225252151489, 'end': 60, 'id': 43867095, 'title':
'https://en.wikipedia.org/wiki/Deadpool_(film)'}]}, {'index': 16, 'utterance': "I haven't
```


Figure 2.3: Example of struggling to disambiguate franchises

The correct annotation should have been, ”Deadpool_2” as Once Upon a Deadpool is a recut version of Deadpool 2. This problem seems to reoccur constantly for all modes.

Another issue with some of the systems is it would make the same spots as the in ground truth, but the link would something else entirely. As shown below it correctly spots Dunkirk, but annotates to Christopher Nolan. One of the actors in the movie. This can be seen in figure 2.4


```
about it?', 'response': []}, {'index': 24, 'utterance': 'Do you like the movie Dunkirk?',
'response': [{'spot': 'Dunkirk', 'start': 22, 'link_probability': 0.5501881837844849, 'rho':
0.3280899226665497, 'end': 29, 'id': 177840, 'title': 'https://en.wikipedia
.org/wiki/Christopher_Nolan'}]}, {'index': 25, 'utterance': 'I have not watched that movie.',
```


Figure 2.4: Correct spot, wrong disambiguation

This proves that there are improvements to be made in the disambiguation part of the entity linking pipeline, there however also problems with the mention detection in some parts. In some cases the systems would not even make spots of entities in the data showing there also improvements to be made on this portion of the pipeline.

2.2 Knowledge Bases

A knowledge base often abbreviated to KB is a machine-readable resource that can contain both structured and unstructured information to be used by computer systems. They are usually more useful when containing structured information, and the ideal representation is called an ontology. This ideal model is structured in a way were it is not just tables containing data, but also pointers to other objects. These objects can in turn have further pointers. These types of knowledge bases are extremely useful as they do not only provide information about data points themselves, but also information about the relationship between the data. This is well illustrated in figure 2.5. Nirenburg and Mahesh (1997)

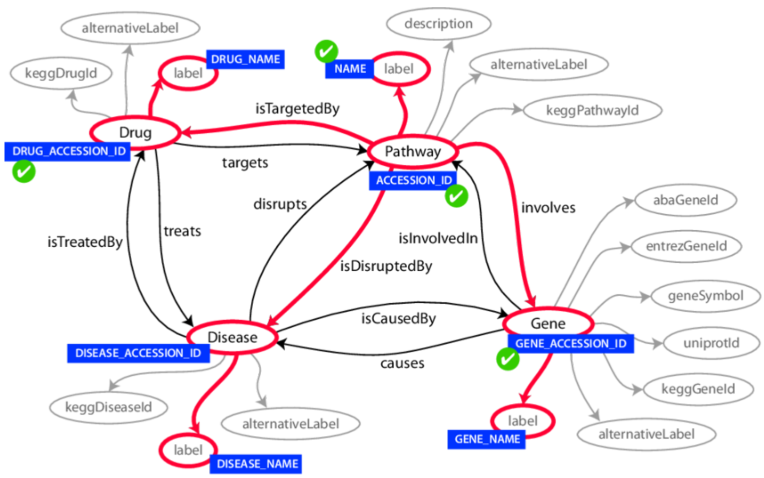

Figure 2.5: Ontology, example taken from Ont

Knowledge bases are as the name might imply are often used in knowledge-based systems. Since many of these systems use inference engines to reason and solve complex problems, they usually need some way to draw these inferences. This is why they usually strictly use the knowledge base models discussed above, ontologies.

The most known and probably biggest knowledge base in the world is wikipedia. It is an online encyclopedia that is driven by volunteer effort. With 6.1 million articles, the english wikipedia is huge catalog of knowledge. And while each article contains some type of structure in the form of title, anchor text, summary etc, this is not what we talk about when we say structured information. There does however exist a structured verson of wikipedia in the form of dbpedia. This knowledge base uses Resource Description Framework or RDF for short to represent extracted information and their relations. Nirenburg and Mahesh (1997)Lehmann et al. (2015)

2.2.1 Semantic Web

Semantic web is a proposed extension of the World Wide Web, the goal is to make the internet machine-readable. The proposed solution for achieving this is adding metadata to otherwise existing content and data through technologies proposed by the World Wide Web Consortium(W3C). The basic concept is notched on Linked Open Data(LOD), a structured data representation modeled as a knowledge graph. LOD includes the actual data about specific entities and concepts, as well as the ontologies. Ontologies shortly described being the classes of objects, their attributes and relationship types between entities.

As mentioned the way researcher and volunteers are trying to achieve is through the standardisation of technologies. One of the these proposed technologies by W3C is the earlier mentioned RDF, but there is also SPARQL, OWL and SKOS. All of these technologies serve a purpose. RDF is the one allowing for the linking of data, OWL is used to build ontologies(or vocabularies as their also called) and SPARQL allows you to actually query this data in any meaningful way. Berners-Lee et al. (2001)Goos et al. (2011)

2.3 Named Entity Recognition and Disambiguation

2.3.1 Information Extraction

Information extraction (IE) as a field in computer science is often thought of as sitting between information retrieval and NLP, with some overlaps. The goal of information extraction is the automated retrieval of structured information such as entities, relationships between entities, and attributes describing entities from usually unstructured documents, but they can also be structured or semi-structured. One such way of achieving structure is the concept of semantic triples, which are statements structured in the form of subject - predicate - object. As already mentioned IE often crosses path with NLP as most cases involve the processing of natural human languages, and is widely used in as as subtask in problems such as Question Answering Systems, Named Entity Linking, Relation Extraction etc. Mahic (2019)Cowie and Wilks (1996)Tang et al. (2007).

2.3.2 Named Entity Recognition

Named Entity Recognition (NER) as the name might suggest the automated extraction or identification of named entities. Entities can be the names of peoples, places, companies, locations and more. NER is probably the first step in many IE tasks, and in essence works by taking an input text and returns all the found named entities. After this is done one can also take it one step further and then classify the entity into a pre-defined category. This is sometimes also included under the NER umbrella, but is usually referred to as Named Entity Recognition and Classification. There are several approaches to NER where some of the most popular are a lexicon approach where you rely on a knowledge base called an ontology, a rule-based system that emply a series of grammatical rules formed by linguists beforehand, a machine-learning based systems that learn from previous examples they have seen or some ensemble or hybrid approach of the aforementioned approaches. Mahic (2019) Powley and Dale (2007) Ritter et al. (2011).

2.3.3 Named Entity Disambiguation

Named Entity Disambiguation (NED) represents the task of disambiguating entities within a text, and aims to link these to the correct entry in some kind of knowledge base, for example Wikipedia, yago, dbpedia. This is a problematic task because some entities can be highly ambiguous and can link to several different entries in the knowledge base

To illustrate we can imagine the utterance, ”Tottenham is one of the poorer areas of London.” Linking London to the correct entry should be pretty straightforward, it is clearly talking about the city and should correspond to that entry in the knowledge base. The problem in the example above lies at Tottenham, for a human it would relatively uncomplicated to draw inference some inference from the rest of the utterance and conclude it is an area in London. This is a bit more problematic for a machine however, when doing a google search for Tottenham the top hit is Tottenham Hotspurs, a football club native to the north London are of Tottenham. Naively just linking to the top hit of this given query would then link the entity to the wrong entry. This is why we need methods to resolves disputes in cases with ambiguous entities.

Based on what kind entity linking system we use there are different approaches to this problem. For a text-based approach a ranking algorithm is usually used as the second step. But ranking the possible nodes is not necessary an easy task either, it is hard to select scoring algorithms, possibly making several nodes a possible match for the given entity. There are other ways to find the best candidates. For instance, one can use some kind of machine learning approach such as SVMs, decision tress, supervised learning, etc. In modern entity linking systems however a graph based approach is used instead. These types of systems employ a large knowledge graph created from a knowledge base, and then take advantage of the graphs topology through complex features to rank a subgraph of the knowledge base. Usual algorithms used here are Pagerank or HITS.

Another common problem is the variety, due to several ways of writing named entities can have many different surface forms in texts. An imagined example could be, ”Chelsea come back to win 4-3 after being down 3 goals to nil at half time. The Blues continue their winning record at home.” In this example Chelsea and the Blues both refer to the same named entity, ”Chelsea F.C.” As shown an entity might have many types of alias surfaces such as nicknames, abbreviations, acronyms, etc. It is problematic matching all the surface forms to corresponding entries.

We also distinguish entity linking systems into two distinct approaches, End-to-end and Disambiguation-Only. For an end-to-end system you process a piece of text to extract entities and then disambiguate these entities to the correct entities in a knowledge base. For disambiguation-only however we take gold standard named entities as input only disambiguate them to the correct entries. Balog (2018).

2.4 Named Entity Linking Pipeline

The most common architecture of a named entity linking system consists of 3 parts.

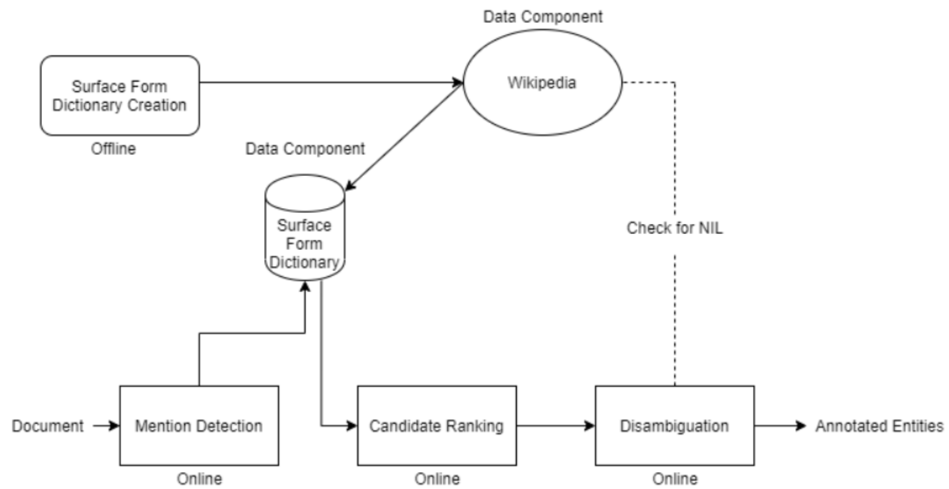

Figure 2.6: A depiction of a simple pipeline

2.4.1 Mention Detection

Before you can being linking mentions to corresponding entries, it is necessary to actually find the possible mentions in a text. This has a few challenges, namely not missing entities that should be linked, finding name variants and filtering out inappropriate ones. This is commonly done by building a dictionary of entity surface forms and then checking all document n-grams against the dictionary. You may filter out undesired entities during this part of the pipeline or late.

2.4.2 Candidate Selection

Next in the pipeline is trying to narrow down the amount of disambiguation possibilities. You create a set or ranked list of candidates for each mention. A costly approach is comparing each mention with all entries. A basic way is to rank the candidates based on their overall popularity. Some other conventional approaches include constructing a bag-of-words representation for each entity that can then be ranked using standard document retrieval techniques. Ideally there should be a good balance between recall and precision, as you do not want to miss the correct entity while also have a smaller set for the disambiguation part.

2.4.3 Disambiguation

At the end of the pipeline you disambiguate down to one single best candidate or none. This can be approached as a ranking problem and you might just pluck out the top ranked candidate from the previous step if you used a ranked list, but you can also consider additional types of features as clues such as prior importance, contextual similarity and coherence. You can also perform some type of pruning, removing low confidence or semantically meaningless annotations. Alternatively, disambiguation may be approached as an inference problem, with the objective of optimizing the coherence among all entity linking decisions in the document. In most cases we end up with a numerical approach, where we can calculate the distance between mentions and candidate entities, either by heuristics or machine learning.

2.5 AIDA

AIDA performs collective disambiguation using a graph-based approach. The graph is constructed with mentions and their candidate entities as nodes. It uses 2 types of edges,

- mention-entity edges which are between mentions and their candidate entities with weights that capture the similarity between the context of a mention and a candidate
- entity-entity edges which are between different entities with weights that capture the coherence (semantic relatedness) between two entities.

AIDA reduces this graph to a dense subgraph that contains all mention nodes and exactly one mention-entity edge for each mention. Density here refers to the total weight of the sub-graph's edges, or alternatively, to the minimum weighted degree in the sub-graph. To find this dense sub-graph where each mention node is connected to one and only one candidate entity node, AIDA uses a greedy algorithm to compute the sub-graph. In each iteration, it performs two steps:

1. identify the entity node that has the lowest weighted degree (sum of the weights of the node's incident edges), and 2. remove this node and its incident edges from the graph unless it is the last remaining candidate entity for one of the mentions. Hoffart et al. (2011)

Pseudocode for this can be seen in figure 2.7.

2.5.1 Pre-Processing

Almost all dense-subgraph problems are almost inevitably NP-hard as they generalize the Steiner-tree problem, an exact algorithm would be infeasible. To solve this problem an approximation algorithm(Sozio and Gionis (2010)) is used. Because this can lead to incoherent entity-mention mappings a constraint is set, each mention node needs to remain connected to at least one entity. This constraint however may lead to suboptimal results and is why a pre-processing step is used. This is solved by pruning entities that are only remotely related to the mention nodes. For each entity node, AIDA computes the distance from the set of all mention nodes in terms of the sum of the corresponding squared shortest path distances. It then restricts the input graph to the entity nodes that are closest to the mentions. An experimentally determined good choice for the size of this set is five times the number of the mention nodes. Then the iterative greedy method is run on this smaller subgraph. Hoffart et al. (2011)

```
Input: weighted graph G of mentions and entities
Output: result graph with one edge per mention
/* pre-processing phase                                              */
foreach entity node e do
    dist_e ← sum of (weighted) shortest paths to each mention
end
keep entities ℰ_c with lowest dist_e, drop the others
/* main loop                                                         */
objective ← min_{e∈ℰ_c} wd(e)/|ℰ_c|
while G has non-taboo entity do
    /* entity is taboo if last candidate for any mention             */
    e ← non-taboo entity with lowest wd(e)
    ℰ_c ← ℰ_c \ e
    remove e with all its incident edges from G
    mwd ← min_{e∈ℰ_c} wd(e) / |ℰ_c|
    if mwd > objective then
        solution ← G
        objective ← mwd
    end
end
/* post-processing phase                                             */
if feasible then
    process solution by enumerating all possible mention-entity pairs
else
    process solution by local search
end
```

Figure 2.7: Pseudocode of how AIDA works, example taken from Balog (2018)

2.5.2 Post-Processing

The final solution, which maximizes the sum of edge weights, is selected in a post-processing phase. If the graph is sufficiently small, it is feasible to exhaustively consider all possible mention-entity pairs. Otherwise, a faster local (hill-climbing) search algorithm may be used. Hoffart et al. (2011)

2.5.3 Measures

Entity-Entity Coherence: AIDA estimates the semantic relatedness between entities using the Wikipeida link structure. The more frequent two entities co-occur in Wikipedia, the higher their semantic relatedness score should be. Therefore, AIDA estimates the entity-entity coherence using the inlink overlap by the approach refined by Milne and Witten that takes into account the total number N of entities as follows:

$$WLM(e_1, e_2) = 1 - \frac{log(max(|L_e1|, |L_e2|) - log(|L_e1 \cap L_e2|))}{log(|\epsilon|) - log(min(|L_e1|, |L_e2|))}$$

Mention-Entity Similarity: For the Mention to Entity AIDA computes a specificity weight for each word that occurs in a keyphrase. This is denoted as the MI for mutual information, and is calculated through joint probabilites as follows:

$$p(e, w) = \frac{|w \in (KP(e) \cup U_{e' \in NI_e} KP(e'))|}{N}$$

with e being entity and w keyword and N denotes the total number of entities. Since there is a chance keyphrases only turn up partially in an input text yet another measure is used for these partial matches. This is done by matching individual words and rewarding their nearness by taking the shortest window of words that contain a maximal number of words of the keyphrase, and is done as follows:

$$score(q) = z(\frac{\sum_{w \in cover} weight(w)}{\sum_{w \in q} weight(w)})^2$$

where z =

$$\frac{matchingwords}{lengthofcover(q)}$$

and weight(w) is either the MI weight defined above or the collection-wide IDF weight. The final simalarity of a mention m to candidate e is aggregated over all keyphrases of e and partial matches giving us the score,

$$simscore(m, e) = \sum_{q \in KP(e)} score(q)$$

2.6 Systems

2.6.1 DBPedia Spotlight

As explained in the DBPedia FAQ DBP (a), ”DBpedia Spotlight is a tool for automatically annotating mentions of DBpedia resources in text, providing a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia.” It works It does this using two distinct and different approaches, Model and Lucene. As outlined in Daiber et al. (2013) for the Model approach, and Mendes et al. (2011) for Lucene, they are both probabilistic approaches and rely on context. And as further outlined a more verbose text makes the disambiguation process more efficient. It works in 4 steps DBP (b). First it does spotting through the identification of surface form substrings of original input. Then it goes through candidate selection where it picks out a set of surface forms from step 1 along with their corresponding DBpedia resource link. Thirdly it does the disambiguation deciding on the most likely candidate, and lastly it does filtering. Where it adjusts the annotations to some requirement(s) provided by the user. The web service that is freely available online uses the newer approach, Model. This approach has been shown to have better performance across the line than the Lucene.

2.6.2 TagMe

In contrast to DBPedia Spotlight, which is partly made to annotate normal length text, tagme was designed with the goal of annotating very shorts texts, like tweets.

Another difference from dbpedia is that tagme uses a three stage process, parsing, disambiguation and pruning. ”Spots” are detected in the text by searching for multi-word sequences in an anchor dictionary. These anchors are text in a wikipedia article that are used as a link to another article. Along with the anchor dictionary, a page catalog is made which the disambiguation step uses. It cross-references anchors detected to relevant pages in the page catalog. The pruning step may then get rid of annotations it does not consider meaningful. Meaning is measured by a scoring function that uses both the link probability of an anchor and measures coherence from a prospective annotation. Ferragina and Scaiella (2010) Mahic (2019)

2.6.3 Google Cloud Natural Language

Unfortunately, since the google cloud is proprietary, the implementation of the natural language service is in a black box. There is no good way to say how it works, or what kind of techniques google uses.

Chapter 3

Approach

3.1 Annotation Process

As the purpose of this thesis was to test entity linking services in a conversational setting and a continuation of the pre-project, the CCPE-dataset made by Google was naturally picked over a more traditional dataset for entity linking like the Wiki-Annot30 dataset, or the TAC KBP English Entity Linking Comprehensive and Evaluation Data 2010 dataset. The CCPE-dataset consists of 502 english dialogs and 12000 annotated utterances of two people discussing movie preferences. The average amount of utterances per conversation is 23,84, with and average of 1,27 entities per utterance. The two main entity types are ”Movie_Genre_Or_Category” and ”Movie_OR_Series”, but there is also the occasional ”Person” type. The data was made using a Wizard-of-Oz approach where two people follow the CCPE method to elicit movie preferences.

The CCPE-dataset is structured as a json file. Consisting of a list of 502 conversations uniquely identified through the conversationId field. Each conversation also has an utterances field containing a list of all utterances for that conversation. Each utterance has an index, speaker(who could be either user or assistant) and a list of annotated segments.

Unfortunately this dataset was only annotated with categorize’s, and did not have target entries in a KB annotated. To be of any use for this thesis these needed to be added. A two step approach was used,

1. Run a script adding disambiguations from wikipedia based on entity names in the data.
2. Manually go through each disambiguation and correct.

The script runs a wikipedia search through their API and returns a list of likely pages from a search word, the search word being an entity name. This list is stored decrementally, with the first item being the most likely. The script would add the first link for entities to the dataset. However this is a naive approach the as the most likely hit does not have to be the correct one. It was mostly done to speed up the process of annotating. An example of this naivety giving us a wrongly annotated link can be seen in the picture below,


```
"index": 14,
"segments": [
  {
    "annotations": [
      {
        "WIKI_LINK": "https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber_of_Secrets",
        "annotationType": "ENTITY_NAME",
        "entityType": "MOVIE_OR_SERIES"
      }
    ],
    "endIndex": 49,
    "startIndex": 10,
    "text": "Harry Potter and the Chamber of Secrets"
  }
],
"speaker": "ASSISTANT",
"text": "how about Harry Potter and the Chamber of Secrets"
```


The second part of the process was to establish an actual ground truth. A human manually adding or correcting annotations has been the gold standard for datasets used in entity linking tasks and was the process used for this dataset and thesis as well. An example of this correction can be seen in the highlighted part of the picture below,

```
"index": 14,
"segments": [
  {
    "annotations": [
      {
        "WIKI_LINK": "https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber_of_Secrets_(film)",
        "annotationType": "ENTITY_NAME",
        "entityType": "MOVIE_OR_SERIES"
      }
    ],
    "endIndex": 49,
    "startIndex": 10,
    "text": "Harry Potter and the Chamber of Secrets"
  }
],
"speaker": "ASSISTANT",
"text": "how about Harry Potter and the Chamber of Secrets"
```


"_(film)" is added to the end in comparison to our last example

You can read further of the TAC-KBP under the evaluation chapter. For the pre-project 50 of the conversations were manually annotated and corrected. As this too small a sample size all 502 conversations were corrected for the thesis. Mahic (2019)

3.2 Initial

This section describes the initial planned approach.

The initial idea was to implement a simple version of AIDA to gather some initial results. These results would be the baseline for any comparisons later on.

This version of AIDA would still use the Milne and Witten approach for the entity-entity coherence but simplify the scoring approach for measuring mention to entity similarity. A combination of a simple contextual feature like cosine similarity and a simple context independent feature like commonness was thought to suffice.

AIDA has a pre-processing step as described in 2.5.1, this step was thought of as unnecessary for this simple AIDA. As explained the dataset is a conversation between two people following a Wizard-Of-Oz approach. As only a single utterance would be evaluated at any time the maximum amount of mention nodes would be rather constricted and small. As a solution in case we got many, a naive approach like only picking the top 10 entity nodes by rank from mention-to-similarity measure and cutting the rest.

The work was going smoothly until the need for the entity to entity scoring algorithm. As mentioned in section 2.5.3 the approach AIDA uses for entity coherence is the Milne and Witten method. As explained in the theory section this approach gives a score based on the overlap in backlinks. The scoring function would be easy enough to implement, but the impasse occurred at creating a reverse-index off all the entries and their backlinks. Due to hardware constraints this was not possible both in regards to speed of actually filling up the database and the final database size in terms of hard drive space.

3.3 Final

As the idea of implementing my own version of AIDA proved unfeasible, I needed some other way to reach the project goals. As developing a novel improvement for a non-existent system makes little sense, the novel improvement would have to be simulated. The idea for the improvement was to help with clarification through the use if follow-up questions and benchmark how this would impact AIDA. To actually check how this simulation would fare we would need some baseline results.

The gathering of these results was done by sending requests to the AIDA JSON web service and gathering the replies. These were then later evaluated against the ground truth annotated dataset for the baseline precision, recall and f1 measures.

Simulating was handled by handpicking conversations and instances were AIDA seemed to struggle and simulating both a system response question and answer. Then checking how the propagation of this new information would impact the measures.

Chapter 4

Evaluation

4.1 Evaluation Methods

In this section I will introduce the common measures for evaluating the performance of entity linking. As mentioned in the pre-project some project teams implementing entity linking systems decide to also invent and develop their own evaluation method and data sets. This poses a problem as biases can occur, and it is possible to design an evaluation method and data set to fit their system perfectly. This is unfortunate and you might run into a situation where your evaluation method gives you a good score, but it ends up performing considerably worse for any other method and data set. This is why no ”designer” metrics were chosen. Rosales-Méndez (2019) Mahic (2019).

As also mentioned in the pre-project one of the fairer ways to evaluate is developed by TAC-KBP Heng Ji and Florian (2015), a conference that specializes in Natural Language Processing tasks. Their evaluation method calculates precision and recall between what they define as the gold standard(G)(which is links annotated manually by a human), and a system's (S) annotations) The annotations are a set of distinct tuples. Values for precision (P) and recall (R) are combined as their balanced harmonic mean (F1), which is used to compare each system. Some other ways to evaluate entity linking systems are Gerbil Micro-F1 and Macro-F1 for End-to-end approaches and Micro-Precision and Macro-Precision for Disambiguation-only approaches. nlp nam Mahic (2019)

The actual gathering of the ”baseline” results was done through API calls to Max Planck’s Institute of Informatics AIDA JSON web service. This API does not allow any tweaking of parameters.

As previously mentioned the dataset is divided into conversations, each with their own list of utterances. The measure scores were set for each conversation. This was done by,

- Looping through the list of utterances and gathering each response into a responses list.
- Cleaning each response for only the necessary data(disambiguated entities, offsets, length and metadata)
- Matching disambiguated entities in the responses to their corresponding gold truth entity by matching utterance index, offset in the utterance and length

After having gathered the baseline results, the impact of clarification was done through simulation. The simulation works by checking how a clarification question might impact the entity-entity and mention-entity measures.

4.2 Evaluation Measures

The equation for regular precision, recall and f1 is,

$$P = \frac{TP}{TP + FP} \tag{4.1}$$

$$R = \frac{TP}{Total} \tag{4.2}$$

$$F1 = \frac{P * R}{P + R} \tag{4.3}$$

Were TP is true positives, FP is false positives and total which is the sum of true positives and false negatives.

Macro precision and recall is just the mean,

$$P_\mu = \frac{P_1 + \cdots + P_n}{n} \tag{4.4}$$

$$R_\mu = \frac{R_1 + \cdots + R_n}{n} \tag{4.5}$$

Where P_μ is the mean from P_1, the precision score from result 1, up to P_n, the precision score of result n. It is the same for R_μ, which is the mean from result R_1 up to result P_n. Macro averaged F1 is the mean F1 across all F1 scores. The micro average precision and recall is however calculated by the following equations,

$$P_M = \frac{TP_1 + \cdots + TP_n}{TP_1 + \cdots + TP_n + FP_1 + \cdots + FP_n} \tag{4.6}$$

$$R_M = \frac{TP_1 + \cdots + TP_n}{T_1 + \cdots + T_n} \tag{4.7}$$

P_M is calculated by using TP and FP from result 1 up to result n. While R_M is calculated using TP and Total from result 1 up to result n. The micro averaged F1 score is calculated as F1 using the micro average precision and recall scores. Mahic (2019)

4.3 Results

4.3.1 Exception Handling

While gathering the results some rare occurrences of Division by Zero errors would occur. These would coincide with the annotator not returning any annotations, giving 0 true positives and 0 false positives. One could handle this by assigning 0 values to precision, recall and F1, but this a naive way of handling these errors. This would however be naive, as in cases where the document actually has 0 entities to annotate, the annotator returning 0 annotations would be correct behavior. This was therefore handled by assigning 1 if there was nothing to annotate and the annotator did not return anything, meaning 0 false positives. There are no such occurrences in this dataset, but one should be aware if you were to try another.

4.3.2 Baseline

As noted earlier these baseline results were gathered by disambiguating utterences from each conversation. This was done through a python scrip looping through the utterences for each conversation in the dataset and gathering the responses.

The baseline results for the 100 first conversations through the python script can be seen below,

Micro		Macro	
Precision	0.348	Precision	0.284
Recall	0.075	Recall	0.061
F1	0.123	F1	0.096

Table 4.1: Table of Micro and Macro averaged precision, recall and F1

One of the immediate eye catchers is the recall metric. In comparison to the entity linking systems benchmarked in the pre-study it seems low. Two tables for comparison are added below. Mahic (2019)

Micro measures,

[]

Micro	Single Utterance		Previous and Current		Full Context	
DBPedia Spotlight	Precision	0.394	Precision	0.384	Precision	0.360
	Recall	0.355	Recall	0.348	Recall	0.329
	F1	0.373	F1	0.365	F1	0.344
TagMe	Precision	0.337	Precision	0.372	Precision	0.387
	Recall	0.414	Recall	0.338	Recall	0.309
	F1	0.371	F1	0.354	F1	0.344
Google Cloud	Precision	0.620	Precision	0.608	Precision	0.607
	Recall	0.265	Recall	0.262	Recall	0.258
	F1	0.371	F1	0.367	F1	0.362

Table 4.2: Micro averaged results from the pre-project

And macro measures,

Macro	Single Utterance		Previous and Current		Full Context	
DBPedia Spotlight	Precision	0.388	Precision	0.381	Precision	0.343
	Recall	0.358	Recall	0.351	Recall	0.327
	F1	0.356	F1	0.350	F1	0.321
TagMe	Precision	0.347	Precision	0.398	Precision	0.415
	Recall	0.404	Recall	0.315	Recall	0.304
	F1	0.373	F1	0.351	F1	0.351
Google Cloud	Precision	0.575	Precision	0.578	Precision	0.572
	Recall	0.262	Recall	0.260	Recall	0.251
	F1	0.360	F1	0.359	F1	0.349

Table 4.3: Macro average results from the pre-project

[]

From a quick glance even the worst performing system in the recall measure had a three time bigger score. The low recall is hard to explain, but one simple reason might be the spotter used not being well suited for this dataset. Another more specific reason might be size of the database AIDA uses. It uses a YAGO2 knowledge with nearly 3 million named entities, quite a bit fewer than the 6.1 million articles on wikipedia, or 17 million YAGO3 has.

4.3.3 Simulation

The simulations will by done by handpicking conversations and utterances AIDA seemed to struggle with. We will also have to assume part of the reason for the issues is related to the smaller database and act as if the entities exist.

The first simulation will be on conversation 15 with conversationId CCPE-55417.

```
Linked entity: https://en.wikipedia.org/wiki/Logan, Ohio
Ground Truth: https://en.wikipedia.org/wiki/Logan_(film)
Results for conversation: 15
Precision: 0.0 Recall: 0.0 F1: 0
```

Figure 4.1: Results for conversation CCPE-55417

From all the utterances in this conversation AIDA had one correct spot in line with the ground truth on utterance index 18 with startindex in text 13. This was also the only spot AIDA made for this conversation. A deeper look at the utterance that gave us this response shows us that problem here seems to be that the mention candidate is wrong.

{'originalFileName': None, 'originalText': 'Yes, I have. Logan is a great movie.', 'overallTime': '149', 'allEntities': ['YAGO:Logan\\u002c_Ohio'], 'entityMetadata': {'YAGO:Logan\\u002c_Ohio': {'knowledgebase': 'YAGO', 'depictionurl': 'http://upload.wikimedia.org/wikipedia/commons/9/9d/Logan_Ohio.jpg', 'depictionthumbnailurl': 'http://upload.wikimedia.org/wikipedia/commons/thumbLogan_Ohio.jpg/200px-Logan_Ohio.jpg', 'importance': 0.04027585384693164, 'entityId': 'Logan\\u002c_Ohio', 'type': ['YAGO_wordnet_district_108552138', 'YAGO_yagoPermanentlyLocatedEntity', 'YAGO_wikicategory_Cities_in_Hocking_County\\u002c_Ohio', 'YAGO_yagoLegalActorGeo', 'YAGO_wordnet_urban_area_108675967', 'YAGO_wikicategory_County_seats_in_Ohio', 'YAGO_wordnet_municipality_108626283', 'YAGO_wordnet_center_108523483', 'YAGO_wikicategory_Cities_in_Ohio', 'YAGO_wordnet_seat_108647945', 'YAGO_wordnet_region_108630985', 'YAGO_yagoGeoEntity', 'YAGO_wordnet_physical_entity_100001930', 'YAGO_wordnet_site_108651247', 'YAGO_wordnet_entity_100001740', 'YAGO_wordnet_object_100002684', 'YAGO_wordnet_area_108497294', 'YAGO_wordnet_city_108524735', 'YAGO_wordnet_county_seat_108547143', 'YAGO_wordnet_tract_108673395', 'YAGO_wordnet_geographical_area_108574314', 'YAGO_wikicategory_Populated_places_established_in_1816', 'YAGO_wordnet_location_100027167', 'YAGO_wordnet_administrative_district_108491826'], 'readableRepr': 'Logan, Ohio', 'url': 'http://en.wikipedia.org/wiki/Logan%2C%20Ohio'}}, 'annotatedText': 'Yes, I have. [[YAGO:Logan\\u002c_Ohio|Logan]] is a great movie.', 'docID': '635775275_1598177017483', 'mentions': [{'allEntities': [{'kbIdentifier': 'YAGO:Logan\\u002c_Ohio', 'disambiguationScore': '0.14617'}], 'offset': 13, 'name': 'Logan', 'length': 5, 'bestEntity': {'kbIdentifier': 'YAGO:Logan\\u002c_Ohio', 'disambiguationScore': '0.14617'}}], 'formatVersion': '2.3', 'allTypes': ['YAGO_wikicategory_Cities_in_Hocking_County\\u002c_Ohio', 'YAGO_wordnet_municipality_108626283', 'YAGO_yagoGeoEntity', 'YAGO_wordnet_region_108630985', 'YAGO_wordnet_geographical_area_108574314', 'YAGO_yagoLegalActorGeo', 'YAGO_wordnet_tract_108673395', 'YAGO_wikicategory_Populated_places_established_in_1816', 'YAGO_wordnet_location_100027167', 'YAGO_wordnet_center_108523483', 'YAGO_wordnet_city_108524735', 'YAGO_wordnet_seat_108647945', 'YAGO_wordnet_physical_entity_100001930', 'YAGO_wordnet_site_108651247', 'YAGO_yagoPermanentlyLocatedEntity', 'YAGO_wordnet_object_100002684', 'YAGO_wordnet_district_108552138', 'YAGO_wordnet_entity_100001740', 'YAGO_wordnet_administrative_district_108491826', 'YAGO_wordnet_county_seat_108547143', 'YAGO_wordnet_urban_area_108675967', 'YAGO_wikicategory_Cities_in_Ohio', 'YAGO_wikicategory_County_seats_in_Ohio', 'YAGO_wordnet_area_108497294']}

Figure 4.2: AIDA JSON Web Service Response

The mention candidate is

'kbIdentifier': 'YAGO:Logan\\u002c_Ohio'

This might because the YAGO2 database used is from 2017, and the movie also came out in 2017. It just might not have been added in the KB.

In any case the wikipedia disambiguation page gives us plenty entity candidates for a knowledge graph. We can simulate this with mention, "Logan" and candidate entities from the disambiguation page. The mention-entity similarity scores were measured by using the online API from twinword.

The initial graph with 3 entities to the mention "Logan" can be seen in figure 4.3.

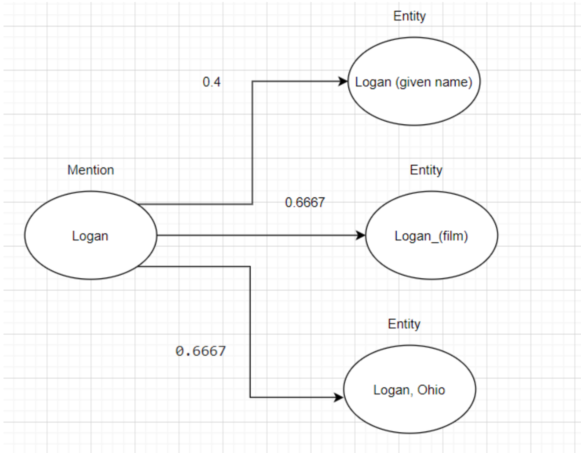

Figure 4.3: Initial graph with calculated mention-entity scores

In the initial graph there is no need for clarification here, the basic AIDA algorithm can run and remove the node with the lowest weighted degree. This can be demonstrated in figure 4.4

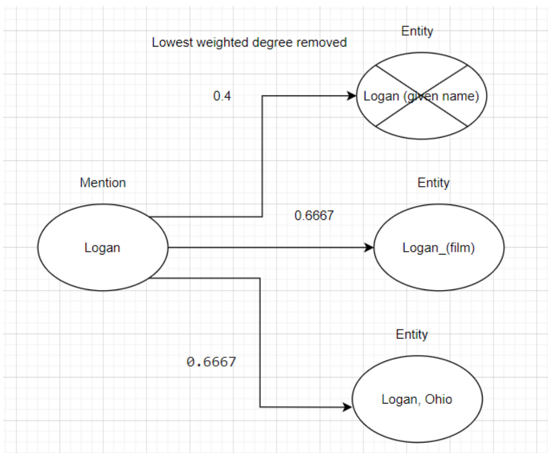

Figure 4.4: Removing node with lowest weighted degree

After having removed the node we now face an issue in that the two nodes that are left have the same mention to entity score. We need some clarification.

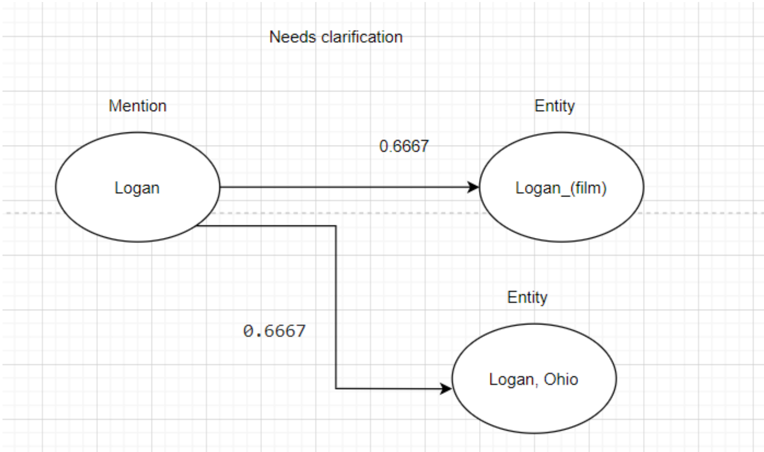

Figure 4.5: Two nodes with same weighted degree

The utterances this knowledge graph was built from was,

```
Yes, I have. Logan is a great movie.
```

The imaginary systems picks up on the context and creates a follow up question. The answer creates a new mention. The imaginary system is shown in figure 4.6

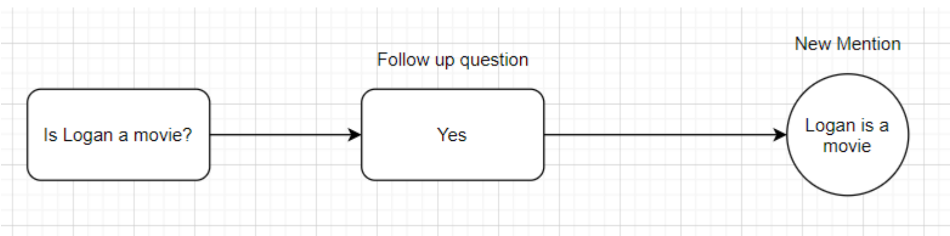

Figure 4.6: Follow-up question

With this new mention we need to recalculate the mention to entity distance. The new calculations are shown in figure 4.7

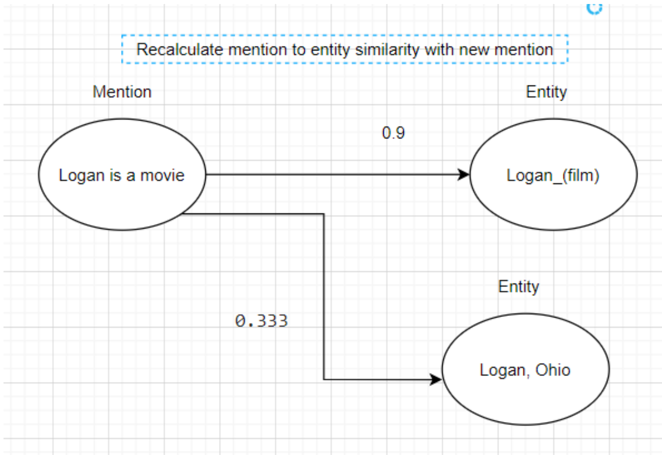

Figure 4.7: Graph with new mention and recalculated scores

After having recalculated there is no more ambiguity in the knowledge graph and we can proceed with removing the lowest weighted degree. Shown in figure 4.8

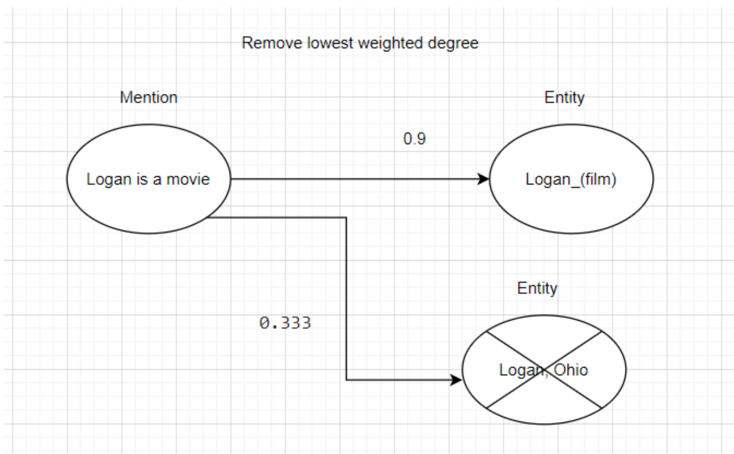

Figure 4.8: Continue with algorithm and remove node with lowest weighted degree

The lowest weighted node has been removed and we are left with one entity for the mention. This entity is now also the same as in the ground truth. This depicted in figure 4.9

This process was repeated for several conversations were appropriate and yielded similar

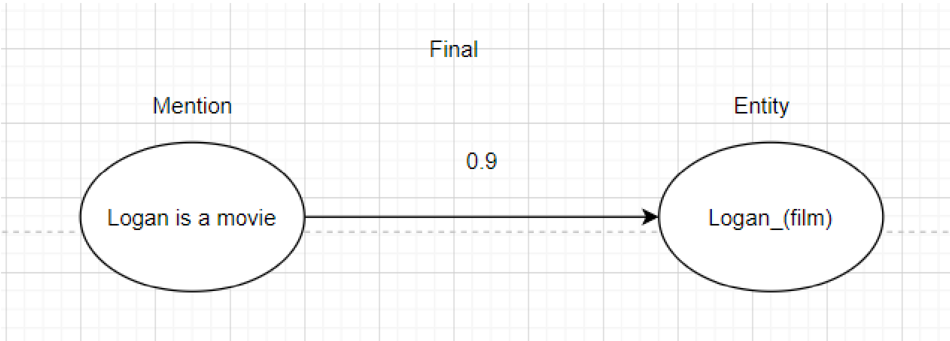

Figure 4.9: Final graph with 1 entity for each mention(In this case only 1 mention)

results. Some of the other tested conversations were conversationId CCPE-70fc4 with utterance 8 and 9, conversationId CCPE-f8c9f with utterance 17. For all the instances this process was tried it correctly disambiguated all but 1, utterance 15 in conversation CCPE-8e113. It ended up disambiguating to the first deadpool film, while the deadpool 2 is correct.

4.4 Analysis

From the spots in the 100 first conversations this imaginary system was tested on, it would on the surface seem to have potential. This potential though seems to really show in events where you can differentiate the named entities into categories, e.g. films, comics and books. This is however as long as you can capture the semantic relatedness between the mention and the entities as the twinword API does.

Where the imaginary system might struggle however is in the same way some of the benchmarked systems in the preproject(Mahic (2019)) did. These are when the semantic relatedness of surface forms and their corresponding entries in the KB are hard to capture. As was the case in the preproject this was often in regards to sequels of movies, or getting the correct entry of a franchise.

As already mentioned in the results, AIDA performed poorly compared to the other entity linking systems in regards to recall. Also as mentioned this might in large part be due to the ”small” amount of entities in the postgres database AIDA uses. This simplifies a lot of the processes, but since entities might not exist as entries in the database this might impact recall by not finding candidates, but might also have impacted the precision by AIDA just not having the correct entry to disambiguate to.

There might be some concerns on the simulations only being done on knowledge graphs containing one mention, but since the imaginary follow-up questions only target 1 mention at a time the new calculations would only impact the score between the target mention and it's entity links. The entity to entity scores would remain the same, as well as the other remaining mention to entity scores.

However there are some concerns in regards to how to structure these clarification questions. One would need to capture the context in some meaningful way so that the answer actually strengthens the link between one or several of the entities connected to the target mention.

For this study the dataset is about movie preferences, and it is easy for a human to construct a question based on the context, this might not however be so easy for a machine. Furthermore since most of the gold truth entities in the dataset are movies, they often contain the suffix ”_(film)”. If the imaginary system is able to capture the context and structure a question like, ”Is this a movie?”, the answer yes and the new mention ”entity is a movie” provides semantic relatedness between ”movie” and ”_(film” parts. This might not be the case for all entities. This would have to explored further by testing with other datasets.

Chapter 5

Conclusions

As mentioned introductory in this thesis the task of named entity linking is not trivial. Be this because of the ambiguity of entities, or the variety of surface forms in text. This problem is even further complicated when in a conversational setting. As the preliminary study for this thesis showed that the systems benchmarked performed inadequately when in a context-poor environment.

On the objectives set forth at the start of this thesis, we did recognize the main errors of the earlier mentioned systems through the error analysis. One of the main struggles was disambiguating to the correct category. The systems would often make the correct spots, but then end up disambiguating to the character or comic instead of the film. Or in some cases disambiguate to the film instead of the whole franchise. There were some other issues, but this was the most prevalent and made the most sense to correct.

To go about this we set out to implement a baseline entity linking system from which we could gather baseline results to be used for the improvement. Since this was cut short due to hardware constraints we ended up simulating an imaginary system.

These simulations gave us some answer on the impact of clarification questions in case of uncertainty. In cases where the named entity can easily be categorized and we can then gauge semantic relatedness between mention and entity in a meaningful way these questions gave the correct ground truth disambiguation in all but one occurrence. In the one case where it failed, it seemed to be due to one of the same issues the systems from the preliminary study had. It struggled to capture the ground truth from the surface form of a movie sequel, it would instead disambiguate to the first movie in the franchise.

Even though clarification questions seem to have some potential there is some further work needed to be done. The dataset used for this thesis might not have been ideal to generalize from as it’s all about movie preferences. Most of the clarification questions ended up being, ”Is this a movie”. And as most movie entities end on the ”_(film)” suffix it is easy to capture the semantic relatedness. This might not be the case for other types of entities and needs to be further explored.

Bibliography

, a. Dbpedia spotlight. https://www.dbpedia-spotlight.org/faq. Accessed: 2019-12-6.

, b. Dbpedia spotlight. https://www.dbpedia-spotlight.org. Accessed: 2019-12-6.

, . Nlp progress. http://nlpprogress.com/english/entity_linking.html. Accessed: 2019-11-29.

, . Nlp progress. http://www.davidsbatista.net/blog/2018/05/09/Named_Entity_Evaluation/. Accessed: 2019-11-30.

, . W3 rdf viz. https://www.w3.org/2018/09/rdf-data-viz/.

Balog, K., 2018. Entity Linking. Springer International Publishing, Cham. pp. 147–188. URL: https://doi.org/10.1007/978-3-319-93935-3_5, doi:10.1007/978-3-319-93935-3_5.

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web" in scientific american.

Cowie, J., Wilks, Y., 1996. Information extraction.

Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N., 2013. Improving efficiency and accuracy in multilingual entity extraction, in: Proceedings of the 9th International Conference on Semantic Systems (I-Semantics).

Ferragina, P., Scaiella, U., 2010. Tagme: On-the-fly annotation of short text fragments (by wikipedia entities), pp. 1625–1628. doi:10.1145/1871437.1871689.

Goos, G., Hartmanis, J., Leeuwen, J., Hutchison, D., Pan, J., Chen, H., Kim, H., Li, J.Z., Wu, Z., Horrocks, I., Mizoguchi, R., Wu, Z., 2011. The semantic web, in: Lecture Notes in Computer Science.

Heng Ji, Joel Nothman, B.H., Florian, R., 2015. Tri-lingual entity discovery and linking.

Hoffart, J., Yosef, M.A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., Weikum, G., 2011. Robust disambiguation of named entities in text, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Edinburgh, Scotland, UK.. pp. 782–792. URL: https://www.aclweb.org/anthology/D11-1072.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C., 2015. Dbpedia - a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web 6, 167–195.

Mahic, K., 2019. On the use of context in a conversational setting.

Mendes, P.N., Jakob, M., Garcia-Silva, A., Bizer, C., 2011. Dbpedia spotlight: Shedding light on the web of documents, in: Proceedings of the 7th International Conference on Semantic Systems (I-Semantics).

Nirenburg, S., Mahesh, K., 1997. Knowledge-based systems for natural language processing, in: The Computer Science and Engineering Handbook.

Powley, B., Dale, R., 2007. High accuracy citation extraction and named entity recognition for a heterogeneous corpus of academic papers, pp. 119 – 124. doi:10.1109/NLPKE.2007.4368021.

Ritter, A., Clark, S., Mausam, Etzioni, O., 2011. Named entity recognition in tweets: An experimental study, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Edinburgh, Scotland, UK.. pp. 1524–1534. URL: https://www.aclweb.org/anthology/D11-1141.

Rosales-Méndez, H., 2019. Towards better entity linking evaluation, pp. 50–55. doi:10.1145/3308560.3314199.

Sozio, M., Gionis, A., 2010. The community-search problem and how to plan a successful cocktail party, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 939–948. URL: https://doi.org/10.1145/1835804.1835923, doi:10.1145/1835804.1835923.

Tang, J., Hong, M., Zhang, D., Liang, B., Li, J., 2007. Information extraction: Methodologies and applications. Emerging Technologies of Text Mining: Techniques and Applications doi:10.4018/978-1-59904-373-9.ch001.

Appendix

Python script:

```
import requests
import json
import functools, operator, collections

url = "https://gate.d5.mpi-inf.mpg.de/aida/service/disambiguate"

conversations = []

with open(r'dataWithLinks.json') as f:
    data = json.load(f)

for obj in data:
    conversations.append(obj)


def request(text):
    reqText = {'text': "{}".format(text)}
    req = json.loads(requests.post(url, data=reqText).content.decode('ut

    return req

def cleaner(responses):
    numOfResponses = 0
    prefix = "https://en.wikipedia.org/wiki/"
    for res in responses:
        entry = {'entities': [], 'offset': [], 'length': []}
        for obj in res['response']['mentions']:
            if obj['allEntities']:
```

```
                numOfResponses = numOfResponses + len(obj['allEntities']
                for ent in obj['allEntities']:
                    ent['kbIdentifier'] = prefix + ent['kbIdentifier']
                    entry['entities'].append(ent['kbIdentifier'].replace
                                                .replace(r"\u0028", "(")
                                                .replace(r"\u0029", ")")
                                                .replace(r"\u0027", "%27")
                                                .replace(r"\u002d", "-")
                                                .replace(r"\u0021", "!")
                                                .replace(r"\u002c", ",")
                                                .replace(r"\u0026", "%26")
                                                .replace(r'\u002e', "."))

                    entry['offset'].append(obj['offset'])
                    entry['length'].append(obj['length'])
        res['response'] = entry
    return responses, numOfResponses

def linker(conv):
    responses = []
    for i in range(len(conv['utterances'])):
        responses.append({'index': conv['utterances'][i]['index'],'utter
    return responses

#print(cleaner(linker(conversations[81])))

def evaluate(responses, annotations, numOfResponses):
    hit = 0
    tot_ents = 0
    for i in range(len(annotations)):
        try:
            for anno in annotations[i]['entities']:
                for ent in anno['annotations']:
                    if ent['annotationType'] == 'ENTITY_NAME':
                        tot_ents = tot_ents + 1
                        for obj in responses:
                            if obj['index'] == annotations[i]['index']:
                                for j, resEnt in enumerate(obj['response
                                    if anno['startIndex'] == obj['respon
                                        print("Linked entity: " + obj['r
                                        print("Ground Truth: " + ent['WI
                                        if obj['response']['entities'][j
                                            print("Hit")
                                            hit = hit + 1
        except:
```

```
            continue

    miss = numOfResponses - hit
    return hit, miss, tot_ents

    pass

def run():
    annotations = []
    micro = []
    eval = []
    print("Running:")
    epochs = 102 #16 annotated conversations
    for j in range(epochs):
        for i in conversations[j]['utterances']:
            try:
                annotations.append({'index': i['index'], 'utterance': i[
            except KeyError:
                annotations.append({'index': i['index'], 'utterance': i[

        responses, numOfResponses = cleaner(linker(conversations[j]))
        tp, fp, total = evaluate(responses, annotations, numOfResponses)
        try:
            precision = tp / (tp + fp)
        except ZeroDivisionError:
            precision = 1 if total == 0 and fp == 0 and tp == 0 else 0
        recall = 1 if total == 0 and fp == 0 and tp == 0  else tp / tota
        try:
            f1 = (2 * precision * recall) / (precision + recall)
        except ZeroDivisionError:
            f1 = 0
        micro.append({'tp': tp, 'fp': fp, 'total': total})
        eval.append({'precision': precision, 'recall': recall, 'f1': f1}
        print("Results for conversation: " + str(j))
        print("Precision: " + str(eval[j]['precision']) + " Recall: " +
            eval[j]['f1']))
        annotations = []

    result = dict(functools.reduce(operator.add,
                                   map(collections.Counter, eval)))
    micro = dict(functools.reduce(operator.add,
                                  map(collections.Counter, micro)))

    micro_avg_precision = micro['tp'] / (micro['tp'] + micro['fp'])
    micro_avg_recall = micro['tp'] / micro['total']
```

```
    try:
        micro_avg_f1 = (2 * micro_avg_precision * micro_avg_recall) / (m
    except:
        micro_avg_f1 = 0
    print("Macro results averaged for all conversations:")
    print("Precision: " + str(result['precision'] / epochs) + " Recall: 
        result['recall'] / epochs) + " F1: " + str(result['f1'] / epochs
    print("Micro results averaged for all conversation:")
    print("Precision: " + str(micro_avg_precision) + " Recall: " + str(m

run()
#print(cleaner(linker(conversations[0])))
#print(request(text))
##Kom til CCPE-41cdd, linje 50454

def getEpochs(id):
    epochs = 0
    for conv in conversations:
        if conv['conversationId'] == "CCPE-41cdd":
            return epochs
        epochs = epochs + 1

#print(conversations[81]['conversationId'])
```

Code output from script:

```
    Running:
Linked entity: https://en.wikipedia.org/wiki/Tom_Hanks
Ground Truth: https://en.wikipedia.org/wiki/Tom_Hanks
Hit
Linked entity: https://en.wikipedia.org/wiki/Leonardo_DiCaprio
Ground Truth: https://en.wikipedia.org/wiki/Leonardo_DiCaprio
Hit
Results for conversation: 0
Precision: 0.6666666666666666 Recall: 0.14285714285714285 F1: 0.23529411
Linked entity: https://en.wikipedia.org/wiki/Adam_Sandler
Ground Truth: https://en.wikipedia.org/wiki/Adam_Sandler
Hit
Results for conversation: 1
Precision: 1.0 Recall: 0.1 F1: 0.18181818181818182
Linked entity: https://en.wikipedia.org/wiki/Step_Brothers_(film)
```

```
Ground Truth: https://en.wikipedia.org/wiki/Step_Brothers_(film)
Hit
Results for conversation: 2
Precision: 1.0 Recall: 0.16666666666666666 F1: 0.2857142857142857
Linked entity: https://en.wikipedia.org/wiki/Emily_Rose_(actress)
Ground Truth: https://en.wikipedia.org/wiki/The_Exorcism_of_Emily_Rose
Results for conversation: 3
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Channing_Tatum
Ground Truth: https://en.wikipedia.org/wiki/Channing_Tatum
Hit
Linked entity: https://en.wikipedia.org/wiki/Jamie_Curtis
Ground Truth: https://en.wikipedia.org/wiki/Jamie_Lee_Curtis
Results for conversation: 4
Precision: 0.3333333333333333 Recall: 0.06666666666666667 F1: 0.11111111
Results for conversation: 5
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/African_American
Ground Truth: https://en.wikipedia.org/wiki/African_Americans
Results for conversation: 6
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/The_First_Wives_Club
Ground Truth: https://en.wikipedia.org/wiki/The_First_Wives_Club
Hit
Results for conversation: 7
Precision: 0.5 Recall: 0.16666666666666666 F1: 0.25
Results for conversation: 8
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Jim_Carrey
Ground Truth: https://en.wikipedia.org/wiki/Jim_Carrey
Hit
Results for conversation: 9
Precision: 1.0 Recall: 0.16666666666666666 F1: 0.2857142857142857
Results for conversation: 10
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Disney_Channel
Ground Truth: https://en.wikipedia.org/wiki/List_of_Walt_Disney_Pictures
Results for conversation: 11
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Thor_Halvorssen_Mendoza
Ground Truth: https://en.wikipedia.org/wiki/Thor_(film)
Linked entity: https://en.wikipedia.org/wiki/Thor_Longus
Ground Truth: https://en.wikipedia.org/wiki/Thor_(film)
Results for conversation: 12
Precision: 0.0 Recall: 0.0 F1: 0
```

```
Linked entity: https://en.wikipedia.org/wiki/Richard_Gere
Ground Truth: https://en.wikipedia.org/wiki/Richard_Gere
Hit
Linked entity: https://en.wikipedia.org/wiki/Julia_Roberts
Ground Truth: https://en.wikipedia.org/wiki/Julia_Roberts
Hit
Linked entity: https://en.wikipedia.org/wiki/Jennifer_Garner
Ground Truth: https://en.wikipedia.org/wiki/Jennifer_Garner
Hit
Results for conversation: 13
Precision: 1.0 Recall: 0.17647058823529413 F1: 0.3
Results for conversation: 14
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Logan,_Ohio
Ground Truth: https://en.wikipedia.org/wiki/Logan_(film)
Results for conversation: 15
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Results for conversation: 16
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Jason_Bourne
Ground Truth: https://en.wikipedia.org/wiki/Jason_Bourne
Hit
Results for conversation: 17
Precision: 1.0 Recall: 0.1 F1: 0.18181818181818182
Linked entity: https://en.wikipedia.org/wiki/Adam_Sandler
Ground Truth: https://en.wikipedia.org/wiki/Adam_Sandler
Hit
Linked entity: https://en.wikipedia.org/wiki/Adam_Sandler
Ground Truth: https://en.wikipedia.org/wiki/Adam_Sandler
Hit
Results for conversation: 18
Precision: 0.6666666666666666 Recall: 0.3333333333333333 F1: 0.444444444
Linked entity: https://en.wikipedia.org/wiki/Ender_Wiggin
Ground Truth: https://en.wikipedia.org/wiki/Ender%27s_Game_(film)
Results for conversation: 19
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 20
Precision: 0 Recall: 0.0 F1: 0
Results for conversation: 21
Precision: 0 Recall: 0.0 F1: 0
Results for conversation: 22
```

```
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 23
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Mamma_Mia!
Ground Truth: https://en.wikipedia.org/wiki/Mamma_Mia!_(film)
Results for conversation: 24
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/White_Chicks
Ground Truth: https://en.wikipedia.org/wiki/White_Chicks
Hit
Linked entity: https://en.wikipedia.org/wiki/Indiana_Jones_(franchise)
Ground Truth: https://en.wikipedia.org/wiki/Indiana_Jones_and_the_Kingdo
Results for conversation: 25
Precision: 0.5 Recall: 0.125 F1: 0.2
Linked entity: https://en.wikipedia.org/wiki/Con_Air
Ground Truth: https://en.wikipedia.org/wiki/Con_Air
Hit
Linked entity: https://en.wikipedia.org/wiki/Avengers_(comics)
Ground Truth: https://en.wikipedia.org/wiki/Avengers:_Endgame
Linked entity: https://en.wikipedia.org/wiki/Forrest_Gump
Ground Truth: https://en.wikipedia.org/wiki/Forrest_Gump
Hit
Results for conversation: 26
Precision: 0.6666666666666666 Recall: 0.15384615384615385 F1: 0.25
Linked entity: https://en.wikipedia.org/wiki/American_Psycho
Ground Truth: https://en.wikipedia.org/wiki/American_Psycho
Hit
Results for conversation: 27
Precision: 1.0 Recall: 0.125 F1: 0.2222222222222222
Results for conversation: 28
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Ant-Man_(film)
Ground Truth: https://en.wikipedia.org/wiki/Ant-Man_(film)
Hit
Results for conversation: 29
Precision: 0.2 Recall: 0.05263157894736842 F1: 0.08333333333333333
Results for conversation: 30
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Walt_Disney_Records
Ground Truth: https://en.wikipedia.org/wiki/The_Walt_Disney_Company
Results for conversation: 31
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Jennifer_Lopez
Ground Truth: https://en.wikipedia.org/wiki/Jennifer_Lopez
Hit
```

```
Results for conversation: 32
Precision: 0.5 Recall: 0.1 F1: 0.16666666666666669
Results for conversation: 33
Precision: 0 Recall: 0.0 F1: 0
Results for conversation: 34
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Seinfeld
Ground Truth: https://en.wikipedia.org/wiki/Seinfeld
Hit
Results for conversation: 35
Precision: 0.3333333333333333 Recall: 0.1 F1: 0.15384615384615383
Linked entity: https://en.wikipedia.org/wiki/Jake_Gyllenhaal
Ground Truth: https://en.wikipedia.org/wiki/Jake_Gyllenhaal
Hit
Results for conversation: 36
Precision: 0.5 Recall: 0.08333333333333333 F1: 0.14285714285714285
Results for conversation: 37
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Arnold_Schwarzenegger
Ground Truth: https://en.wikipedia.org/wiki/Arnold_Schwarzenegger
Hit
Linked entity: https://en.wikipedia.org/wiki/Rambo_III
Ground Truth: https://en.wikipedia.org/wiki/Rambo_(franchise)
Linked entity: https://en.wikipedia.org/wiki/Arnold_Schwarzenegger
Ground Truth: https://en.wikipedia.org/wiki/Arnold_Schwarzenegger
Hit
Results for conversation: 38
Precision: 0.6666666666666666 Recall: 0.09523809523809523 F1: 0.1666666
Results for conversation: 39
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 40
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Tom_Hanks
Ground Truth: https://en.wikipedia.org/wiki/Tom_Hanks
Hit
Results for conversation: 41
Precision: 1.0 Recall: 0.1 F1: 0.18181818181818182
Results for conversation: 42
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Steven_Seagal
Ground Truth: https://en.wikipedia.org/wiki/Steven_Seagal
Hit
Results for conversation: 43
Precision: 0.5 Recall: 0.2 F1: 0.28571428571428575
Linked entity: https://en.wikipedia.org/wiki/John_Tucker_Must_Die
```

```
Ground Truth: https://en.wikipedia.org/wiki/John_Tucker_Must_Die
Hit
Results for conversation: 44
Precision: 1.0 Recall: 0.08333333333333333 F1: 0.15384615384615385
Results for conversation: 45
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Paul_Rudd
Ground Truth: https://en.wikipedia.org/wiki/Paul_Rudd
Hit
Results for conversation: 46
Precision: 1.0 Recall: 0.1 F1: 0.18181818181818182
Results for conversation: 47
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Billy_Madison
Ground Truth: https://en.wikipedia.org/wiki/Billy_Madison
Hit
Linked entity: https://en.wikipedia.org/wiki/Happy_Gilmore
Ground Truth: https://en.wikipedia.org/wiki/Happy_Gilmore
Hit
Linked entity: https://en.wikipedia.org/wiki/Adam_Sandler
Ground Truth: https://en.wikipedia.org/wiki/Adam_Sandler
Hit
Results for conversation: 48
Precision: 1.0 Recall: 0.17647058823529413 F1: 0.3
Linked entity: https://en.wikipedia.org/wiki/Lincoln_(2012_film)
Ground Truth: https://en.wikipedia.org/wiki/Lincoln_(film)
Results for conversation: 49
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 50
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber
Results for conversation: 51
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Will_Ferrell
Ground Truth: https://en.wikipedia.org/wiki/Will_Ferrell
Hit
Linked entity: https://en.wikipedia.org/wiki/Harold_%26_Kumar
Ground Truth: https://en.wikipedia.org/wiki/Harold_%26_Kumar_Go_to_White
Results for conversation: 52
Precision: 0.3333333333333333 Recall: 0.16666666666666666 F1: 0.2222222
Results for conversation: 53
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Woody_Allen
Ground Truth: https://en.wikipedia.org/wiki/Woody_Allen_filmography
```

```
Linked entity: https://en.wikipedia.org/wiki/Annie_Hall
Ground Truth: https://en.wikipedia.org/wiki/Annie_Hall
Hit
Linked entity: https://en.wikipedia.org/wiki/Woody_Allen
Ground Truth: https://en.wikipedia.org/wiki/Woody_Allen
Hit
Results for conversation: 54
Precision: 1.0 Recall: 0.25 F1: 0.4
Results for conversation: 55
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Tim_Burton
Ground Truth: https://en.wikipedia.org/wiki/Tim_Burton_filmography
Linked entity: https://en.wikipedia.org/wiki/Tim_Burton
Ground Truth: https://en.wikipedia.org/wiki/Tim_Burton
Hit
Linked entity: https://en.wikipedia.org/wiki/Willy_Wonka
Ground Truth: https://en.wikipedia.org/wiki/Willy_Wonka
Hit
Results for conversation: 56
Precision: 0.5 Recall: 0.13333333333333333 F1: 0.2105263157894737
Results for conversation: 57
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Inside_Man
Ground Truth: https://en.wikipedia.org/wiki/Inside_Man
Hit
Linked entity: https://en.wikipedia.org/wiki/Jason_Bourne
Ground Truth: https://en.wikipedia.org/wiki/Jason_Bourne_(film)
Results for conversation: 58
Precision: 0.5 Recall: 0.125 F1: 0.2
Linked entity: https://en.wikipedia.org/wiki/Jason_Bourne
Ground Truth: https://en.wikipedia.org/wiki/Jason_Bourne_(film)
Results for conversation: 59
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Stanley_Kubrick
Ground Truth: https://en.wikipedia.org/wiki/Stanley_Kubrick
Hit
Results for conversation: 60
Precision: 0.3333333333333333 Recall: 0.1 F1: 0.15384615384615383
Linked entity: https://en.wikipedia.org/wiki/Will_Ferrell
Ground Truth: https://en.wikipedia.org/wiki/Will_Ferrell
Hit
Results for conversation: 61
Precision: 1.0 Recall: 0.125 F1: 0.2222222222222222
Results for conversation: 62
Precision: 0 Recall: 0.0 F1: 0
```

```
Linked entity: https://en.wikipedia.org/wiki/Gabby_Logan
Ground Truth: https://en.wikipedia.org/wiki/Logan_(film)
Linked entity: https://en.wikipedia.org/wiki/Hugh_Jackman
Ground Truth: https://en.wikipedia.org/wiki/Hugh_Jackman
Hit
Results for conversation: 63
Precision: 0.3333333333333333 Recall: 0.125 F1: 0.18181818181818182
Linked entity: https://en.wikipedia.org/wiki/The_First_Wives_Club
Ground Truth: https://en.wikipedia.org/wiki/The_First_Wives_Club
Hit
Linked entity: https://en.wikipedia.org/wiki/Julia_Roberts
Ground Truth: https://en.wikipedia.org/wiki/Julia_Roberts
Hit
Linked entity: https://en.wikipedia.org/wiki/Tom_Cruise
Ground Truth: https://en.wikipedia.org/wiki/Tom_Cruise
Hit
Linked entity: https://en.wikipedia.org/wiki/Sandra_Bullock
Ground Truth: https://en.wikipedia.org/wiki/Sandra_Bullock
Hit
Results for conversation: 64
Precision: 1.0 Recall: 0.2222222222222222 F1: 0.3636363636363636
Linked entity: https://en.wikipedia.org/wiki/Deadpool_(video_game)
Ground Truth: https://en.wikipedia.org/wiki/Deadpool_(film)
Results for conversation: 65
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 66
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Thor_(Marvel_Comics)
Ground Truth: https://en.wikipedia.org/wiki/Thor_(Marvel_Comics)
Hit
Results for conversation: 67
Precision: 1.0 Recall: 0.07692307692307693 F1: 0.14285714285714288
Results for conversation: 68
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 69
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Washington,_D.C.
Ground Truth: https://en.wikipedia.org/wiki/Olympus_Has_Fallen
Results for conversation: 70
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Good_Day_(The_Dresden_Dolls
Ground Truth: https://en.wikipedia.org/wiki/A_Good_Day_to_Die_Hard
Linked entity: https://en.wikipedia.org/wiki/Bruce_Willis
Ground Truth: https://en.wikipedia.org/wiki/Bruce_Willis
Hit
```

```
Linked entity: https://en.wikipedia.org/wiki/John_McClane
Ground Truth: https://en.wikipedia.org/wiki/John_McClane
Hit
Linked entity: https://en.wikipedia.org/wiki/Reservoir_Dogs
Ground Truth: https://en.wikipedia.org/wiki/Reservoir_Dogs
Hit
Linked entity: https://en.wikipedia.org/wiki/Sigourney_Weaver
Ground Truth: https://en.wikipedia.org/wiki/Sigourney_Weaver
Results for conversation: 71
Precision: 0.5 Recall: 0.2 F1: 0.28571428571428575
Linked entity: https://en.wikipedia.org/wiki/Kurt_Russell
Ground Truth: https://en.wikipedia.org/wiki/Kurt_Russell
Hit
Linked entity: https://en.wikipedia.org/wiki/Western_(genre)
Ground Truth: https://en.wikipedia.org/wiki/Western_(genre)
Hit
Linked entity: https://en.wikipedia.org/wiki/Lil_Wayne
Ground Truth: https://en.wikipedia.org/wiki/John_Wayne
Results for conversation: 72
Precision: 0.5 Recall: 0.2 F1: 0.28571428571428575
Results for conversation: 73
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Spider-Man
Ground Truth: https://en.wikipedia.org/wiki/Spider-Man
Hit
Results for conversation: 74
Precision: 0.5 Recall: 0.08333333333333333 F1: 0.14285714285714285
Linked entity: https://en.wikipedia.org/wiki/Vincent_Price
Ground Truth: https://en.wikipedia.org/wiki/Vincent_Price
Hit
Results for conversation: 75
Precision: 1.0 Recall: 0.1111111111111111 F1: 0.19999999999999998
Linked entity: https://en.wikipedia.org/wiki/Western_(genre)
Ground Truth: https://en.wikipedia.org/wiki/Western_(genre)
Hit
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Hit
Results for conversation: 76
Precision: 0.5 Recall: 0.08333333333333333 F1: 0.14285714285714285
Linked entity: https://en.wikipedia.org/wiki/Harrison_Ford
Ground Truth: https://en.wikipedia.org/wiki/Harrison_Ford
Hit
Linked entity: https://en.wikipedia.org/wiki/Santa_Monica,_California
Ground Truth: https://en.wikipedia.org/wiki/Ocean%27s_Eleven
```

```
Linked entity: https://en.wikipedia.org/wiki/Billy_Ocean
Ground Truth: https://en.wikipedia.org/wiki/Ocean%27s_Eleven
Linked entity: https://en.wikipedia.org/wiki/Andy_Serkis
Ground Truth: https://en.wikipedia.org/wiki/Andy_Serkis
Hit
Results for conversation: 77
Precision: 0.4 Recall: 0.18181818181818182 F1: 0.25000000000000006
Linked entity: https://en.wikipedia.org/wiki/Denzel_Washington
Ground Truth: https://en.wikipedia.org/wiki/Denzel_Washington
Hit
Linked entity: https://en.wikipedia.org/wiki/Matt_Damon
Ground Truth: https://en.wikipedia.org/wiki/Matt_Damon
Hit
Linked entity: https://en.wikipedia.org/wiki/Jim_Carrey
Ground Truth: https://en.wikipedia.org/wiki/Jim_Carrey
Hit
Results for conversation: 78
Precision: 0.6 Recall: 0.2727272727272727 F1: 0.37499999999999994
Linked entity: https://en.wikipedia.org/wiki/Syfy
Ground Truth: https://en.wikipedia.org/wiki/Science_fiction_film
Results for conversation: 79
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 80
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Netflix
Ground Truth: https://en.wikipedia.org/wiki/Netflix
Hit
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Hit
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Hit
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Hit
Linked entity: https://en.wikipedia.org/wiki/Quentin_Tarantino
Ground Truth: https://en.wikipedia.org/wiki/Quentin_Tarantino
Hit
```

```
Results for conversation: 81
Precision: 1.0 Recall: 0.21739130434782608 F1: 0.3571428571428571
Linked entity: https://en.wikipedia.org/wiki/Santa_Monica,_California
Ground Truth: https://en.wikipedia.org/wiki/Ocean%27s_8
Results for conversation: 82
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 83
Precision: 0 Recall: 0.0 F1: 0
Results for conversation: 84
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber
Results for conversation: 85
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 86
Precision: 0 Recall: 0.0 F1: 0
Results for conversation: 87
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 88
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Robert_Downey,_Jr.
Ground Truth: https://en.wikipedia.org/wiki/Robert_Downey_Jr.
Linked entity: https://en.wikipedia.org/wiki/Jack_Black
Ground Truth: https://en.wikipedia.org/wiki/Jack_Black
Hit
Results for conversation: 89
Precision: 0.5 Recall: 0.1 F1: 0.16666666666666669
Results for conversation: 90
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Paul_Rudd
Ground Truth: https://en.wikipedia.org/wiki/Paul_Rudd
Hit
Linked entity: https://en.wikipedia.org/wiki/Simon_%26_Schuster
Ground Truth: https://en.wikipedia.org/wiki/Aladdin_(1992_Disney_film)
Results for conversation: 91
Precision: 0.25 Recall: 0.1 F1: 0.14285714285714288
Linked entity: https://en.wikipedia.org/wiki/Eric_Von_Schmidt
Ground Truth: https://en.wikipedia.org/wiki/About_Schmidt
Results for conversation: 92
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Keanu_Reeves
Ground Truth: https://en.wikipedia.org/wiki/Keanu_Reeves
Hit
Results for conversation: 93
Precision: 0.16666666666666666 Recall: 0.09090909090909091 F1: 0.117647
```

```
Linked entity: https://en.wikipedia.org/wiki/Guardians_of_the_Galaxy_(20
Ground Truth: https://en.wikipedia.org/wiki/Guardians_of_the_Galaxy_(fil
Linked entity: https://en.wikipedia.org/wiki/J.U.S.T.I.C.E._League
Ground Truth: https://en.wikipedia.org/wiki/Justice_League_(film)
Linked entity: https://en.wikipedia.org/wiki/Spider-Man_(Miles_Morales)
Ground Truth: https://en.wikipedia.org/wiki/Spider-Man:_Into_the_Spider-
Linked entity: https://en.wikipedia.org/wiki/The_Dark_Knight_Returns
Ground Truth: https://en.wikipedia.org/wiki/Batman
Linked entity: https://en.wikipedia.org/wiki/Captain_America
Ground Truth: https://en.wikipedia.org/wiki/Captain_America:_The_First_A
Linked entity: https://en.wikipedia.org/wiki/Andrew_Garfield
Ground Truth: https://en.wikipedia.org/wiki/Andrew_Garfield
Hit
Linked entity: https://en.wikipedia.org/wiki/Gwen_Stacy
Ground Truth: https://en.wikipedia.org/wiki/Gwen_Stacy
Linked entity: https://en.wikipedia.org/wiki/Mary_Jane_Watson
Ground Truth: https://en.wikipedia.org/wiki/Mary_Jane_Watson
Hit
Results for conversation: 94
Precision: 0.25 Recall: 0.09090909090909091 F1: 0.13333333333333333
Linked entity: https://en.wikipedia.org/wiki/The_Bourne_Ultimatum_(film)
Ground Truth: https://en.wikipedia.org/wiki/The_Bourne_Ultimatum_(film)
Hit
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_and_the_Half-Bl
Results for conversation: 95
Precision: 0.1111111111111111 Recall: 0.14285714285714285 F1: 0.125
Results for conversation: 96
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Gone_Girl_(album)
Ground Truth: https://en.wikipedia.org/wiki/Gone_Girl_(film)
Linked entity: https://en.wikipedia.org/wiki/Harry_Porter
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Linked entity: https://en.wikipedia.org/wiki/Thor_(singer)
Ground Truth: https://en.wikipedia.org/wiki/Thor_(film)
Results for conversation: 97
Precision: 0.0 Recall: 0.0 F1: 0
Results for conversation: 98
Precision: 0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter
Hit
Linked entity: https://en.wikipedia.org/wiki/Country_music
Ground Truth: https://en.wikipedia.org/wiki/Western_(genre)
Linked entity: https://en.wikipedia.org/wiki/Santa_Monica,_California
```

```
Ground Truth: https://en.wikipedia.org/wiki/Ocean%27s_Eleven
Results for conversation: 99
Precision: 0.3333333333333333 Recall: 0.0625 F1: 0.10526315789473684
Linked entity: https://en.wikipedia.org/wiki/Dunkirk_(film)
Ground Truth: https://en.wikipedia.org/wiki/Dunkirk_(2017_film)
Linked entity: https://en.wikipedia.org/wiki/Harry_Potter_(film_series)
Ground Truth: https://en.wikipedia.org/wiki/Harry_Potter_and_the_Chamber
Results for conversation: 100
Precision: 0.0 Recall: 0.0 F1: 0
Linked entity: https://en.wikipedia.org/wiki/Captain_America
Ground Truth: https://en.wikipedia.org/wiki/Captain_America:_The_First_A
Linked entity: https://en.wikipedia.org/wiki/Iron_Man_(song)
Ground Truth: https://en.wikipedia.org/wiki/Iron_Man_(2008_film)
Linked entity: https://en.wikipedia.org/wiki/Iron_Man_(2008_film)
Ground Truth: https://en.wikipedia.org/wiki/Iron_Man_(2008_film)
Hit
Results for conversation: 101
Precision: 0.3333333333333333 Recall: 0.125 F1: 0.18181818181818182
Macro results averaged for all conversations:
Precision: 0.28409586056644875 Recall: 0.06176682327958525 F1: 0.096801
Micro results averaged for all conversation:
Precision: 0.3480392156862745 Recall: 0.07529162248144221 F1: 0.1238012

Process finished with exit code 0
```

NTNU

Kunnskap for en bedre verden