Natural Language Processing (COS4861)

Adriaan Louw (53031377)

October 14, 2018

1   Question 1

Wordnet has 4 noun senses and 3 verb senses for the word "scrap". Noun senses:

Sense 1: a small fragment of something broken off from a whole

Sense 2: worthless material that is to be disposed of

Sense 3: a small piece of something that is left over after the rest has been used

Sense 4: the act of fighting; any contest or struggle

Verb senses:

Sense 5: dispose of something useless or old

Sense 6: have a disagreement over something

Sense 7: make into scrap or refuse

The senses $scrap^1$, $scrap^2$ and $scrap^3$ are related to sense $scrap^4$ by homonymy.

Sense $scrap^4$ is related to sense $scrap^6$ by polysemy because the idea of having a disagreement and fighting is semantically connected. Otherwise stated, disagreements are a common cause for fighting and also disagreeing with someone can be seen as a form of fighting.

The sense $scrap^2$ is related to sense $scrap^7$ via polysemy. In that $scrap^2$ denotes what is created when the action sense $scrap^7$ is done.

2   Question 2

Simulated Verb senses

Sense 1: reproduce someone's behavior or looks

Sense 2: create a representation or model of

Sense 3: make a pretence of

Simulated Adjective senses:

Sense 4: (not genuine or real; being an imitation of the genuine article

Sense 5: reproduced or made to resemble; imitative in character

Synthetic Noun senses

Sense 1: a compound made artificially by chemical reactions

Synthetic Adjective senses:

Sense 2: not of natural origin; prepared or made artificially

Sense 3: involving or of the nature of synthesis (combining separate elements to form a coherent whole) as opposed to analysis

Sense 4: systematic combining of root and modifying elements into single words

Sense 5: of a proposition whose truth value is determined by observation or facts

Sense 6: artificial as if portrayed in a film

Sense 7: not genuine or natural

      Artificial Adjective senses

Sense 1: contrived by art rather than nature

Sense 2: artificially formal

Sense 3: not arising from natural growth or characterized by vital processes

2.0.1 Similarities

Senses simulated[4] and synthetic[7] are similar in that they relate something that is not natural or genuine. Senses synthetic[2] and artificial[3] both denote something that is not of natural origin. Simulated[5] and artificial[2] have the meaning of someone being who they are not.

2.0.2 Differences

Only one of these have a sense that is a noun. That is synthetic[1]. This noun can be described by artificial[3]. Simulated is the only lexeme that has verb senses. Only synthetic has a sense (synthetic[3]) that relates to bringing this together. In other words synthesis.

3 Question 3

3.1 Question 3.1.1

The birr is the currency of Ethiopia. I was using www.google.com as my search engine
    Precision would be:

$$
\begin{aligned}
Precision &= \frac{\mid R \mid}{\mid T \mid} \\
&= \frac{6}{10} \\
&= 0.6
\end{aligned}
\tag{1}
$$

    Recall would be:

$$
\begin{aligned}
Recall &= \frac{\mid R \mid}{\mid U \mid} \\
&= \frac{6}{26} \\
&= 0.23
\end{aligned}
\tag{2}
$$

    A large source of errors would be that manu currency conversion websites came up. From these websites it is not always obvious from which country a specific currency came from. Correct answers were drowned out by all the currency conversion websites.

3.2 Question 3.1.2

Antnio Mascarenhas Monteiro preceded Pedro Pires as the president of Cape Verde. I was using www.google.com as my search engine
    Precision would be:

$$Precision = \frac{|R|}{|T|}$$
$$= \frac{3}{10} \tag{3}$$
$$= 0.3$$

Recall would be:

$$Recall = \frac{|R|}{|U|}$$
$$= \frac{3}{6} \tag{4}$$
$$= 0.5$$

The search engine was focussing on all the websites containing Pedro Peres and was not capable of distinguishing that the important part of the query was who preceded Pedro Peres, not Pedro Peres himself.

3.3 Question 3.2.1

Using www.bing.com as the search engine.

$$Precision = \frac{|R|}{|T|}$$
$$= \frac{5}{10} \tag{5}$$
$$= 0.5$$

Recall would be:

$$Recall = \frac{|R|}{|U|}$$
$$= \frac{5}{21} \tag{6}$$
$$= 0.23$$

The precision of google was higher but the recall was the same. One potential reason for google outperforming bing is that google filtered out many more of the currency exchange websites that were more numerous. Additionally google managed to return a single answer with the word "Ethiopia". Indicating that it is confident enough that that is the answer.

3.4 Question 3.2.2

Using www.bing.com as the search engine

$$Precision = \frac{|R|}{|T|}$$
$$= \frac{3}{10} \tag{7}$$
$$= 0.3$$

Recall would be:

$$Recall = \frac{\mid R \mid}{\mid U \mid}$$
$$= \frac{3}{17} \tag{8}$$
$$= 0.17$$

Bing and google had similar recall. But the precision of google was much higher than bing. This is only because bing managed to find more relevant pages in the first 100 results.

4   Question 4

Mean reciprocal rank (MRR) is a standard used to compute how effective a specific algorithm is at answering a set of questions over a large dataset. The MRR is calculated as follows:

$$MRR = \frac{\sum_{i=1}^{i} \frac{1}{r_i}}{N} \tag{9}$$

where N is the number of questions asked and r_i is the rank of the i'th question.

Each question returns a set of ranked answers. The rank of the first correct answer is then used to calculate the reciprocal of that rank as in the above equation (Jurafsky & Martin, 2014).

References

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). Pearson London.