

Project Internship – Intelligent Agents

CS4514-P – WS2023/24

Exercise 1

Magnus Bender, Ralf Möller

Published: 20.10., 12:00

Presentation: 17.11., 12:15 - 13:45

Getting Started

The purpose of this internship is the implementation of an Information Retrieval (IR) agent based on a collection of Wikipedia articles. The agent answers queries by returning Wikipedia articles from its corpus. Besides answering queries, the agent is also interested in maintaining its corpus, e.g., by extending it with interesting new articles.

We start the internship with the implementation of interactions among agents. In the first exercise, we perform an auction of documents between three agents. In the next exercises, we then improve the methods used by the agents, so that the agents provide better results using methods introduced in the lecture *Web Mining Agents*.

Task 1: Installation and Setup

In the internship, the Smart Python Agent Development Environment (SPADE)¹ framework is used. The first task is to get SPADE running on your computer.

We have created a Docker-Image containing all necessary resources and providing a pre-configured runtime environment. Please refer to the *Setup Guide*² and install Docker as well as Docker-Compose on your computer.

Hint 1 *We strongly recommend to use the Docker-Container provided, however, it is also possible to install the packages on your own and use your own setup.*

In the Moodle course, a *Project Package*³ can be downloaded. The package contains all files needed to start the Docker-Container. Please make sure, that the test agent runs on your computer to complete this task. The output should look like Figure 1.

¹<https://spade-mas.readthedocs.io/en/latest/>

²<https://www.ifis.uni-luebeck.de/~bender/lehre/ws2324/ia/setup/>

³<https://moodle.uni-luebeck.de/mod/resource/view.php?id=421670>

```

State: Preparation (Visitor Otto)
State: Preparation (Visitor Heinz)
State: Preparation (Moin)
State: Moin (Moin)
State: End (Moin)
State: Await Moin, Got message 'Moin Visitor0' (Visitor Otto)
State: Await Moin, Got message 'Moin Visitor1' (Visitor Heinz)
State: END (Visitor Otto)
State: END (Visitor Heinz)

```

Figure 1: The correct output of `test_agent.py`.

Hint 2 Take a look at the test agent consisting of `test_agent.py`, `moin_agent.py` and `visitor_agent.py` in the folder `./project/src/test`.

The Auction

We assume that the agent has a corpus of Wikipedia articles and is interested in extending its corpus with further articles. However, in this exercise, the agent is not able to get every document the agent is interested in, because the agent has to pay for new documents. Therefore, the agent needs to estimate for each offered document the value and has to make the decision of how much money would be appropriate to pay for the document. In this setup, we assume that there is another agent that is also interested in new documents. Thus, this setup results in an auction.

The basic setup is as follows:

- (i) There exist one auctioneer(-agent) and two bidder(-agents).
- (ii) The bidder-agents have given queries which they need to answer and to answer the queries, the agents need to get new, appropriate documents.
- (iii) The auctioneer-agent auctions new documents. The bidder-agents rate the documents value by using `tf.idf` with respect to the given queries.

Task 2: Data Preparation

Both bidder-agents have the same initial corpus of Wikipedia articles. The titles of the articles are given in `corpus.txt`. Additionally, the titles of the documents the auctioneer-agent is selling, are given in `sell.txt`. The `.txt`-files contain titles of Wikipedia articles.

Hint 3 In the `utils` directory of the Project Package is a class `Wikipedia`, which downloads and returns the content of a Wikipedia article given the article's title.

Download the articles to sell and the articles in each agent's corpus. Choose three queries (in form of Wikipedia-articles in the agent's corpus), for each of the two agents. The queries are the different goals of both agents.

Question 1 *Which queries are appropriate for the auction, which are not?*

As the value the agent gives to a specific document is based on its tf.idf-score, agents with the same three query documents will bid the same.

Task 3: Data Preprocessing

The downloaded Wikipedia documents are not yet ready for the usage with tf.idf and need to be preprocessed. Create a function taking text as input and returning an array of words using the Python NLTK⁴ library. The following preprocessing tasks shall be applied:

- (i) Split the articles' content into tokens
- (ii) Convert every token to lowercase
- (iii) Remove punctuation from each token
- (iv) Filter out remaining tokens that are not alphabetic
- (v) Filter out tokens that are stop words
- (vi) Stem words, e.g., with PorterStemmer

Apply your function on all documents, i.e., on all articles to sell and in the corpus.

Task 4: The Auction and the Auctioneer

The auction should be a first-price sealed-bid auction, meaning that each bidder bids independently and sealed (thus, without knowledge of the bidding of the other bidding agent) and the agent which has given the highest bid wins and has to pay its bidding to the auctioneer. Each bidder-agent is thinking about the amount of money it is maximally willing to pay for a document. A rational agent is only willing to pay less than or equal the value the document has for the task of the agent. This means that both bidder-agents could have quite differing opinions of the valuation they are willing to pay for a document. For more information about this auction type, see [RN02].

The auctioneer is a simple structured agent which has a set of documents to sell. The auctioneer-agent sells the documents one after the other to the bidder-agent with the highest bid. For each document, the auctioneer-agent gives the document first to the bidder agents, such that they can evaluate the document and calculate their scoring function (their valuation).

Then the auctioneer gets the bids of the agents and sells the document. The auctioneer only announces the winner of the bidding but does not announce the price the winner is paying and also not the bid of the losing agent.

⁴<https://www.nltk.org/book/>, <https://www.nltk.org/book/ch03.html>

Hint 4 *The auctioneer does not want to maximize its profit, meaning there is no need to implement a specific selling strategy.*

Hint 5 *It seems unintuitive that both agents can look at and calculate with the document before they bought it. Therefore, you can think of the auction process in a way that not the information contained in the document is sold, but the agent is buying the right to publish this information, the right to give this information to the human.*

In a (classical) English auction, the bidders are bidding one after the other until they reach their personal valuation score, thus the maximum they are willing to pay. Therefore, the bidding strategy is a simple step-wise increasing of the bid. Using a first-price sealed-bid auction, this strategy can not be used anymore, as both agents are bidding only one time. Therefore, a more sophisticated bidding strategy is necessary.

In the next task, the agents implement a strategy solely based on their own valuation of a given document.

Task 5: The Bidding Agent – Basic Functionality

The general procedure works as follows: Both agents receive a new document from the auctioneer. Based on this document, the agents calculate the tf.idf-score and, based on this score, a valuation function which assigns the document an amount of money which the agent is willing to pay. Thus, the function shows the utility the document has for the agent.

Then, both agents give a sealed bid for the given document.

Hint 6 *It is not necessary to restrict the amount of money an agent has to be finite, as the value the agent is willing to pay is solely based on the valuation function.*

Think of a valuation function. The valuation function may be based on:

- The tf.idf-score; the assumption is, that the relation between tf.idf-score and valuation is not linear, but above a specific threshold, the value increases, e.g., quadratic.
- The improvement of the query-answering after adding the new document to the corpus compared to answering the question based on the corpus without adding the new document (a small improvement is less valuable than a great improvement, even when the great improvement is based on a lower tf.idf-score).

It is only possible to bid positive values or to give no bid for a document.

Question 2 *Should the agent use only the three query documents to calculate the value of the valuation function? How could an agent also use its corpus?*

Hint 7 *Think of the (unlikely) case that the bids of both agents are the same including no bid.*

Hint 8 *Incorporate the fact that it is possible that a document is valuable for only one or for more than one query.*

- (i) Create a valuation function.
- (ii) Both agents have the same functionality (but different query documents, meaning the bidding will not totally be the same).
- (iii) The agents should bid as much as they give valuation to the given document, meaning the valuation score represents the price to pay.
- (iv) The winning agent may update its tf.idf-model with the bought document.

Question 3 *Is there a better bidding strategy not depending on any anticipation of the other agents behaviour than bidding exactly the valuation?*

Task 6: Evaluation of the Agent's Quality

In this task, the quality of both agents should be evaluated. Therefore, the following functions can be used.

Hint 9 *It is also allowed to define your own evaluation functions.*

The first evaluation function is based on the documents the agent missed because of its bidding strategy and such the valuations the agent missed. On the one hand, all valuations of the documents the agent did buy are summed up. On the other hand, a second sum is created by summing up all valuations which the agent could have bought (sum of all bids). This two sums can then be compared resulting in a quality score.

The second evaluation function is based on the amount the agent overpaid, e.g., paying 20 while the other agent only bids 5 results in a loss of 14, as a bid of 6 would have been sufficient. For all documents the agent bought, the bid of the other agent is considered and the difference (minus one) is summed up as loss.

Question 4 *Why are these scores (the scores you choose, respectively) appropriate to reason about the agent's bidding quality? Is there an improvement for the anticipating agent?*

Outlook

In the next exercises, the bidding strategy will be improved in the way that one agent tries to anticipate the bidding of the other agent and improves its strategy in this way. Further, tf.idf will be accompanied with other methods and answering queries by the agents will be implemented.

References

- [RN02] RUSSELL, S. ; NORVIG, P.: *Artificial Intelligence: A Modern Approach*. Pearson, 2002