

**Project Internship – Intelligent Agents**  
**CS4514-P – WS2023/24**

## **Exercise 3**

Magnus Bender, Ralf Möller

Published: 15.12., 12:00

Presentation: 26.01., 12:15 - 13:45

### **Preface**

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.

In the last two exercises, we have implemented three different types of agents (i.e., auctioneer, IR agent, and questioner). These agents perform an auction of documents from Wikipedia articles and simultaneously run an information retrieval service. The IR agent answers queries based on its corpus and also bids in an auction to extend its corpus.

Currently, the strategy for calculating the bid for a document is very simple and does not take into account the bids and actions of the other agents. In this exercise, we will improve the bidding strategies and adjust the possible queries asked by the questioners.

We apply techniques introduced in the lecture *Web Mining Agents*, especially we build on game theory and the theory of agent collaboration. In the end, the IR agents use an advanced bidding strategy and anticipate the activities of the other agents participating at the auction.

**Hint 1** *We have updated the Project Package in the Moodle course for exercise 3.*

### **Tasks**

This exercise is divided into four tasks. First, we expand the set of possible queries used by the questioners. Second, each bidder has to anticipate the scope of the competing bidder and estimate the competitor's queries during the auction. Third, we define multiple strategies to calculate the bid for a document using the estimated competitor's scope. Finally, we evaluate the performance of the agents and strategies.

## Task 1: Changing the Questioner

Again, in the project package are multiple text files containing the titles of Wikipedia articles. Besides the **corpus.txt** and **sell.txt** there are six files containing queries **queries\_i.txt** to **queries\_vi.txt**.

**Hint 2** *We have updated the articles in all files and randomly shuffled some of them. There exists a file **gold.txt** which contains all articles grouped by topic and divided into queries and others. The articles are roughly from six different topics and each topic has 10 queries, 10 articles in corpus, and 20 articles to sell. There are 7 articles to sell that do not belong to any topic.*

Each of the six sets of queries defines a scope for one questioner. On startup, each questioner randomly chooses one set without laying back. As before, each questioner asks the queries from the set of queries to its IR agent. When anticipating the competitor for the auction, the IR agent has to estimate the set of queriers that the competitor is using. In our setting, there are five possible sets the competing IR agent may have.

**Hint 3** *For each set of queries you might choose different hyperparameters, e.g., for the topic model. However, the same hyperparameters should always be used for the same set of queries.*

**Hint 4** *We selected six sets of queries to reduce the complexity estimating the competitor's scope.*

The questioner shall award an reward for each query answered by the IR agent. The reward is a natural number between 0 and 10, where 0 should indicate incompatible documents, i.e., a *useless* answer, while 10 should indicate multiple compatible documents, i.e., a *useful* answer.

**Question 1** *What is a good strategy to calculate the reward for an answer? What does make an answer subjectively useful?*

---

Checklist Task 1:

- ☐ Questioner uses one of the six sets of queries.
- ☐ Questioner awards rewards to the IR agent.

## Task 2: Anticipate Competitor

Considering the chosen auction type, the strategy used in the exercise 1 is not the optimal one. It is possible that one document has a high valuation for the first IR agent but a rather low one for the second IR agent. Then, the first agent will win the bidding but has payed way too much money. We already saw this in our evaluation of exercise 1.

Therefore, it would be helpful for the first agent to anticipate the valuation the other agent gives to the document. Then, the agent can bid less money and however, can be sure to win the document. To anticipate the valuation we need to know the scope of the competing agent, i.e., the set of queries that the competing agent has to answer. We know all six sets of queries (five possibilities, after removing our own), our own bid and the winner of each auction. Also, we know the basic valuation function for a document (see rules in Task 4) and the initial corpus.

**Question 2** *Imagine the competitor uses a simple strategy and always bids the calculated value of a document. How could one calculate all (five) possible values? Given all values and the information which agent won the auction, what conclusions can be drawn?*

Implement a strategy to estimate the scope of the competing agent. Remember that the agents use some non-deterministic algorithms and thus make sure to gain enough evidence.

**Hint 5** *For the correct anticipation, it is necessary to not only incorporate the knowledge of the agents bidding, but it might also be necessary to incorporate the documents the other agent bought.*

**Hint 6** *It is sufficient to develop a strategy estimating the query set against the simple bidding strategy, only.*

---

Checklist Task 2:

- ☐ IR agent uses a strategy to estimate the competitor's scope.
- ☐ The (current) corpus and most probable queries of the competitor are *known*.

### Task 3: Bidding Strategies

In the last task, we developed a strategy to estimate the queries of the competing IR agent, based on this estimations we can calculate the bid for a document now. The advanced strategies are based on the basic valuation function. Depending on the different strategies, the bid might be below or above the value.

The value of a document for the IR agent itself shall be calculated similar to exercise 1 and 2. This valuation function(s) have to comply with the following rules:

- The basic value is based on the (current) corpus and the queries.
- Each value and each bid must be a natural number between 0 and 100 (0 means no bid and nobody gets the document if all agents bid 0).
- The advanced calculation must only use the estimated queries of the competitor and the queries sent to the agent itself and the (current) corpus. Especially, the IR agent must not use the six sets of queries besides the estimation gained from task 2.

- The simple bidding strategy always bids the value of a document.

The advanced strategies first calculate the basic value and follow a (single) second goal afterwards. There are multiple goals a strategy may follow. Choose and implement three goals from the following. The IR agents should have a parameter to select a strategy.

- (i) There exists a maximum amount of money the agent is allowed to spend. The agent can freely decide how to spend the money. This amount is calculated by the formula

$$75 \cdot \frac{|\{\text{Documents to sell}\}|}{6}.$$

- (ii) The difference of the value of the bought documents and the money spend to buy them must always be positive or zero. If the agents buys a document spending less money than its value, the agent might spend more money for the next document. The simple bidding strategy fulfills this goal, too.
- (iii) The rewards awarded from the questioner are maximized. (It might be helpful to split the queries in two parts, the first part is send at the beginning and the second part send at the end. Thus, the agent can estimate its scope using the first part and get good rewards answering the second part after the auction.)
- (iv) Save as much money as possible while having enough documents. The ratio between money spent and value of the corpus is maximized.
- (v) Imagine your own goal and implement a strategy to reach it.

**Question 3** *Which strategies did you choose and why?*

---

Checklist Task 3:

- ☐ IR agent may run three different bidding strategies.

## Task 4: Extending the Evaluation

We extend the evaluation of the exercise 1 with two additional functions.

The third evaluation function is based on the score usually used for the evaluation of auctions [RN02]. The evaluation function is based on the amount of money the agent saved with its bidding strategy. It is the difference between the amount of money the agent was willing to pay and the amount of money it actually paid. This is summed up over all documents where the agent won the bidding.

The fourth evaluation function is based on the sum of rewards the agent received from its questioner.

**Question 4** *Run the simple agent from exercise 1 against the new agent(s).*

**Question 5** *Run the different agents with different strategies against each other.*

**Question 6** *You might also try to run agents pursuing the same strategy against each other.*

---

Checklist Task 4:

- ☐ All agents print informative information and scores while running.

**Hint 7** *Again, the solution of exercise 3 shall be an improved version of the solution of exercise 2 (i.e., recursively also of exercise 1). The improvement consists of the new bidding strategies and changes in the questioner.*

## References

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