Web Information Retrieval

Assignment 1

Team:

- Alena Zorich <elzorich@uni-koblenz.de>
- Igor Maksimovich Fedotov <ifedotov@uni-koblenz.de>
- Mariya Chkalova <mchkalova@uni-koblenz.de>
- Arsenii Smyrnov <smyrnov@uni-koblenz.de>

1 Machine Learning and its relation to Information Retrieval (14 Points)

---

1. Research on your own and give a definition on what it means for a computer program to learn in the context of Machine Learning.

We show the algorithm a set of data and it analyzes the data, creating and modifying its own internal variables based on the data shown, tries to find some patterns in the data. This is called learning. Then, when we show him other set of data, the algorithm will be able to use its parameters to classify new data.

---

2. Explain the difference between *supervised* and *unsupervised* learning.

With supervised learning we are aware what a correct output should look like, so we have a clue between the relationship of input and output (classification). In contrast, with unsupervised learning there will be no feedback based on the prediction results, since we do not know how should our results look like (clustering, AR mining).

---

3. Explain the difference between flat and hierarchical clustering.

Flat clustering is when we just split data onto separate clusters. When the clustering is hierarchical, smaller cluster get joined into bigger clusters, until whole data is one cluster (or the other way around for top-down hierarchical clustering).

---

4. Research or come up with a scenario in which Machine Learning techniques could be used in Information Retrieval and describe it.

Clustering techniques can be used to group similar documents to improve indexing and then showing more relevant results for users' queries.
Classification of text with naive bayes approach.