Exam 1 - Computation of TF-IDF

The words car, auto, best should have the following frequencies in the ten documents Doc1, .., Doc10. Simply assume that all other words in the documents are stop words.

term \doc	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
car	3	0	0	5	12	0	0	2	8	1
auto	8	6	0	12	0	0	9	1	3	10
best	0	1	7	0	1	5	12	0	2	0

Compute the tf-idf weight for each term using the equations in the slides 14-16.

Exam 2 - Inverted Lists

Assume the following documents

doc-id\text	
Doc1	“italy is world champion 2006”
Doc2	“germany and italy played each other in the semifinal”
Doc3	“germany was in the semifinal 2006”
Doc4	„germany won the semifinal in italy 1990”

Assume that the following terms are stop words: is, and, in, the, was, each, other.
Construct an inverted index according to the figure in slide 25.

Exercises for CLIR and CLQA

- What do we mean when we say a IR system is monolingual, multi-lingual or cross-lingual ?

- In the Machine Translation based approach for cross-lingual IR three different approaches are distinguished. Name them and briefly describe them.

- In the index translation-based approach of CLIR, why is ambiguity of index terms a problem ?

- Name a major difference between cross-lingual IR and cross-lingual QA.

Continued

- For the following questions translate each to English using Google translate, Bing Translator and a online translation system of your choice and merge the resulting three different translations into a single common one using a bag-of-words approach (following the idea of step 1 in slide 49):

  „Welches Land wurde 1978 Fußballweltmeister?“
  „Коя страна стана световен шампион по футбол за 1978 г.?“
  „Mistä maasta tuli jalkapallon maailmanmestari vuonna 1978?“
  „¿Qué país se proclamó campeón en los Mundiales de Fútbol de 1978?“

- Send the translated questions to a search engine. Do this for the individual translations and for the merged translation. Assess the resulting documents as „relevant“ or „non-relevant“.

Exams for Textual Inference

- For the following text pairs (T,H)
  - decide whether entailment holds (indicated by „YES“) ot not (indicated „NO“)
  - check out, whether a simple bag-of-words approach is sufficient (you may consider some lexical preprocessing, like stemming, POS-tagging)
  - if not, check out what additional knowledge sources might be helpful (syntax, lexical semantics, domain or world knowledge, ...)

Example 1

(entailment="???")
<t>Mr. Sherwood speculated that the leeway that Sea Containers has means that Temple would have to "substantially increase their bid if they're going to top us."</t>
<h>Someone would have to increase the bid.</h>

Example 2

(entailment="???")
<t>After all, if you were going to set up a workshop you had to have the proper equipment and that was that.</t>
<h>Somebody had to have the equipment.</h>

Example 3

(entailment="???")
<t>The first was that America had become or was in danger of becoming a second-rate military power.</t>
<h>America was in danger.</h>

Example 4

(entailment="???")
<t>Interpublic Group said its television programming operations which it expanded earlier this year agreed to supply more than 4,000 hours of original programming across Europe in 1990.</t>
<h>Interpublic Group expanded.</h>

Example 5

(entailment="???")
<t>"Oh," said the woman, "I've seen that picture already."</t>
<h>The woman has seen something.</h>

Example 6

(entailment="???")
<t>The Big Board also added computer capacity to handle huge surges in trading volume.</t>
<h>Surges were handled.</h>

Example 7

(entailment="???")
<t>Most of them are Democrats and nearly all consider themselves, and are viewed as, liberals.
</t>
<h>Some consider themselves liberal.</h>

Example 8

- For the following hypthothesis H, a set of texts T_i are given that all entail H.
- Let A = syntactic information, B = lexical semantics, C = inference.
- Assign to each T_i one of A, B, or C, which means that either A or B or C is the major information source to decide entailment. Briefly indicate, on what basis your decision was made.

Example 8

H: People were forced to leave their pets behind when they evacuated New Orleans.

T1: Thousands of people were forced to leave their pets behind when they evacuated New Orleans. (A or B or C ?)

T2: Animal rescue offcials have been collecting scores of pets and other animals from the shattered city, while many survivors have told of their distress at having to leave beloved cats and dogs behind in the watery city when they fled. (A or B or C ?)

T3: Such emotional scenes were repeated perhaps thousands of times along the Gulf Coast last week as pet owners were forced to abandon their animals in the midst of evacuation. (A or B or C ?)

T4: For Elizabeth Finch, the owner of two dogs named Zorra and Hans Blix, the sight of citizens forced to choose between their pets and their safety was, like the disaster itself, indicative of broader social rifts. (A or B or C ?)

T5: The animals are being cared for at a farm north of Louisiana until they can be reunited with their families, many of whom were told they would not be able to bring their pets on evacuation buses and helicopters. (A or B or C ?)