NTCIR-9 1CLICK nugget creation policy

Oct 28, 2010 by Tetsuya Sakai (tesakai@microsoft.com)

General policies:

- A nugget is a short factual statement such that an assessor can judge whether a given text shows or clearly implies that statement to be true.
- Each nugget must be factual. It should not reflect an opinion or a controversial claim.
- Information available on official web pages is considered factual. Information available from other web sources is considered factual provided that they do not contradict with the official information.
- Events that occur in 2011 or after will be ignored. Thus, nuggets are built based on established facts as of 2010.
- Some queries may have multiple possible intents. For these queries, nuggets are built for each intent. At the evaluation stage, nuggets for major intents may be weighted more highly than those for minor intents.

Query-type specific policies:

CELEBRITY queries

- We assume that the user wants to gather various facts about a famous person (actor, politician, artist etc).
- Various facts: date/place of birth, real name, blood type, height, hobbies, profession, personal history, awards, publications, discography, films and TV series in which the person appeared, favourite baseball team, favourite food etc.
- As the output text is supposed to be very short, we will NOT reward inclusion of information such as
    o Appearances in TV commercials
    o Appearances in one-off TV shows
    o Appearances in computer games
  We assume that there are more important pieces of information that should be displayed to the user. We will also exclude information on the person's weight, shoe size and vital statistics.

LOCAL queries

- We assume that the user wants to physically visit or contact a facility (school, shop, office, amusement park, hotel, train station etc).
- Hence we expect systems to return facts such as: postal and email addresses, phone and fax numbers, opening hours, how to access the facility by train/bus/car, nearest stations, time required for the travel, whether the facility has a car park and its opening hours etc.

- We will NOT reward inclusion of money-related information such as entrance fees, parking fees etc.
- As the output text is supposed to be very short, we will NOT reward inclusion of bus/train timetable information.

DEFINITION queries

- We assume that the user is not familiar with a term and wants to look understand its meaning.

QA (question answering) queries

- We assume that the user wants an answer to simple factual questions. Unlike other question types, these queries are given in natural language rather than a single phrase. The nuggets must be part of the answer that satisfies the natural language question. The questions may elicit simple factoids, a series of definitions, and occasionally answers that require some very simple inferences from multiple pieces of information.