**Event Mention Detection scoring**

The scorer reads the output of event mention detection systems and compares them to the gold standard.

**Input:**

Gold standard annotation for a text, in format (one line per mention)

System output annotation for the same text, in format (one line per mention)

**Output:**

System output annotation for the same text, with score appended to each line

**BNF definition of format:**

discontiguous-mention := contiguous-mention + contiguous-mention

contiguous-mention := <engine ID> <text ID> <block ID> <offset within block>

<length of mention> <mention> <event-type> <realis-status>

<engine ID> := the name of the system

<text ID> := the ID of the text

<block ID> := the ID number of the text block

<offset within block> := the starting position (in characters) of the mention string

<length of mention> := the number of characters of the event mention string

<mention> := the actual character string of the mention

<event-type> := the ACE hierarchy type

<realis-status> := the REALIS label

<score1> := any score (confidence, etc.) the system wants to assign (ignored)

<score2> := score assigned in the evaluation

<score3> := additional possible score assigned by human

**Scoring:**

Let SLineNo := next line in System result input file

Let S := the mention’s string (at line==SLineNo in the System file)

Let GLineNo := the line in the Gold Standard file whose mention corresponds to S

(see note 1 below)

Let G := the mention’s string at line==GLineNo in the Gold Standard file

if G == S, then score := 1.0

if G∧S == {}, then score := 0.0

else

let G := G – {invisible-words} (see note 2 below)

let S := S – {invisible-words}

if |G| > |S|, then score := (|S∧G|)/|G|

if |G| < |S|, then score := (|S∧G|)/|S|

append score to the end of SLineNo in the System file, in position <score2>

Note 1: Simply look for the line in the Gold Standard file whose mention includes the first character of the mention in the current SLineNo of the System file.

Note 2: Invisible words are ignored in scoring. They include: determiners, …?

Examples:

Rule 1: do not accept prepositions but include particles

* "[look] up a chimney" vs "[look up] a dictionary"
* "[climb] up the ladder"
* [take responsibility for]
* sing [all the way] to school
* [go] to school

Rule 2: consider the maximum extent of an event mention, but don't worry about determiners (they are invisible)

* [takes a shower] ==> it is okay for annotators to include "a" in their annotation; we can ignore "a" in evaluation
* [make a quick decision] ==> it is okay for annotators to annotate the whole phrase; we can ignore "a" and include "quick" in evaluation