

# Measuring the contribution of Word Sense Disambiguation for QA

---

- Provide English documents and queries (plus Spanish queries) with automatically annotated word senses (WordNet)
  - News corpora only: LA94 GH95
- Sense information added to all content words
  - Lemma
  - Part of speech
  - Weight of each sense in WordNet 1.6
- XML with DTD for documents and queries
- Significant effort (100Mword corpus), shared with the Robust-WSD task

# Measuring the contribution of Word Sense Disambiguation for QA

```
<DOC>
  <DOCNO>GH950102-000000</DOCNO>
  <DOCID>GH950102-000000</DOCID>

  <HEADLINE>
    <TERM ID="GH950102-000000-1" LEMA="alien" POS="JJ">
      <WF>Alien</WF>
      <SYNSET SCORE="0.6" CODE="01295935-a"/>
      <SYNSET SCORE="0.4" CODE="00984080-a"/>
    </TERM>

    <TERM ID="GH950102-000000-2" LEMA="treatment" POS="NN">
      <WF>treatment</WF>
      <SYNSET SCORE="0.827904118008605" CODE="00735486-n"/>
      <SYNSET SCORE="0" CODE="03857483-n"/>
      <SYNSET SCORE="0.172095881991395" CODE="00430183-n"/>
      <SYNSET SCORE="0" CODE="05340429-n"/>
    </TERM>
```

# Measuring the contribution of Word Sense Disambiguation for QA

---

- Participants explore how the word senses (plus the semantic information in wordnets) can be used in QA
- Relevant tasks:
  - ES2ENG and EN2EN (QA<sub>WSD</sub> participants only)
  - 2 runs, with and without WSD
- Reduced interest, one participant only
  - Low performance with and without WSD