

Plan for tagging of sentences (BTP)

Utkarsh Upadhyay

December 9, 2008

Tagging is being done at two stages corresponding to the two stage architecture of the project.

1 Sentence Level

Possible tags are:

- No tags (Useless)
- Tagged as Useful
- Tagged as deceptive

These tags should help one in determining which sentences to pick and which to disregard, helping in fine tuning the summarizer. Also, with the GUI's help, the tagging boundaries are made to coincide with proper constituent boundaries, i.e., the parts selected will be complete sub-trees of the original parse tree.

2 Sub-sentence tags

2.1 Tags

All template related tags will probably look like:

- Predicate
- Arg-TMP
- Arg1 (absolute_value)

- Arg2 (relative_value)
- Arg3 (direction)
- ...

However, at the sentence level, we are also assigning the following tags:

- **Keyword** : These are the non-verb predicates. Predicates automatically qualify for keywords.
- **Named entity**

to help us in making a database of the corresponding.

2.2 Templates

All templates will have these common entries:

- **Name**: Filled by looking for positive anaphoric resolution in neighbouring sentences.
- **Time-frame**: Filled by looking for references in the sentence only or Global information in titles, etc.

The possible templates thought so far are:

- **Profit/Loss**:
 - **Keyword predicates (stemmed)**: profit, loss
 - **Absolute amount**
 - **Relative amount**
 - * Reference to old data/old estimates
 - * Percentage
 - * Absolute increase/decrease
 - * **Direction**
- **Estimates**:
 - **Keyword predicates (stemmed)**: estimate, revenue
 - **Absolute amount**
 - **Relative amount**

- * Reference to old data/old estimates
- * Percentage
- * Absolute increase/decrease
- * **Direction**

- **Share prices:**

- **Keywords (stemmed):** share, target price
- **Absolute Amount**
- **Relative Amount**
 - * Reference to old data/old estimates
 - * Percentage
 - * Absolute increase/decrease
 - * **Direction**

2.3 Current status

I am working on assigning the tags to the template elements using the GUI. The structure in all the templates is easy to see and I am working on exploiting it. A sample tagging is given below:

Shares fell \$3.91, or 9.5 percent, to close at \$37.46.

```
<useful>
  <keyword>Shares</keyword>
  <predicate>fell</predicate>
  <absolute_amount>$3.91</absolute_amount>,
  or
  <relative_amount>9.5 percent</relative_amount>,
  to close at $37.46.
</useful>
```

However, as more sentences are encountered, the number of tags, type of tags, keywords, etc. would grow.