A Public Reference Implementation of the RAP Anaphora Resolution Algorithm

Long Qiu, Min-Yen Kan, Tat-Seng Chua
School of Computing
National University of Singapore
3 Science Drive 2
Singapore 117543
qiull@comp.nus.edu.sg

Abstract
We present a standalone, publicly-available implementation of the Resolution of Anaphora Procedure (RAP) given by Lappin and Lea (1994). The RAP algorithm resolves third person pronouns, lexical anaphors, and identifies pleonastic pronouns. Our implementation, JavaRAP, fills a current need in anaphora resolution research by providing a reference implementation that can be benchmarked against current algorithms. The implementation uses the standard, publicly available Charniak (2000) parser as input, and generates a list of anaphora-antecedent pairs as output. Alternately, an in-place Evaluation on the MUC-6(Message Understanding Conferences) co-reference task shows that JavaRAP has an accuracy of 57.9%.

JavaRAP: A Public Reference RAP Implementation

• What is JavaRAP for?
  – To benchmark other anaphora resolution algorithms;
  – To provide anaphora resolution to NLP applications (automatic text summarization, Q&A, etc.).
• Software required:
  – Java JDK 1.4 +
  – The Charniak parser (www.cs.brown.edu/people/ec/)
• Associated Tools:
  – Sentence Splitter
  – Rule-Based
  – Flexible (supports different input/output formats)
  – Anaphora Resolver Evaluator
  – Pair-wise comparison between resolution result and annotation (MUC6 co-reference convention)
• Availability: Free and downloadable from “www.comp.nus.edu.sg/~qiul/NLPTools”

Resolution of Anaphora Procedure

• Lexical Anaphors: include reflexives (itself, ourselves, etc.) and reciprocals (each other, one another, ...).
• A lexical anaphor A is coreferential with a NP if any of the following conditions holds:
  (a) A is in the argument domain of N and N fills a higher argument slot than A (“They wanted to see themselves.”); or
  (b) A is in the adjunct domain of N (“He worked by himself.”);
  (c) A is in the NP domain of N (“John likes Bill’s portrait of himself.”);
  (d) N is an argument of a verb V; Meanwhile, there is a noun phrase Q in the argument domain or the adjunct domain of N such that Q has no noun determiner, and (i) A is an argument of Q, or (ii) A is an argument of a preposition PREP which is an argument of Q (“They told stories about themselves.”); or
  (e) A is a determiner of a noun Q and (i) Q is in the argument domain of N and N fills a higher argument slot than Q, or (ii) Q is in the adjunct domain of N (“John and Murry like each other’s portraits.”).

Anaphor Binding Algorithm for lexical anaphors

• Syntactic Filter
  – for third personal pronouns

Table 1: Salience Factors and their initial weights (as in RAP)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-adverbial Emphasis</td>
<td>0.4</td>
</tr>
<tr>
<td>Complement Emphasis</td>
<td>0.3</td>
</tr>
<tr>
<td>Accusative Emphasis</td>
<td>0.1</td>
</tr>
<tr>
<td>Existential Emphasis</td>
<td>0.6</td>
</tr>
<tr>
<td>Salience Measure</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Sample Input and Output of JavaRAP

Input:
Eastern Airlines executives notified union leaders that the carrier wishes to discuss selective wage reductions on Feb. 3.

Output:
1. (2,8) Union representatives who could be reached ~ (2,7) they;
   (2,7) they ~ (2,2) they;
2. (1) Eastern Airlines executives notified union leaders that the carrier wishes to discuss selective wage reductions on Feb. 3.

Pleonastic Pronoun Detection

• Pleonastic Pronouns: pronouns that have no referent.
• Detected by pattern matching:
  – It is Model_Adjective that S
  – It is time to VP

Syntactic Filter

• A third person pronoun P is NOT coreferential with a NP if any of the following conditions holds:
  (a) P is in the argument domain of NP;
  (b) P is in the adjunct domain of NP;
  (c) P is in a head H and NP is contained in H;
  (d) P is in the NP domain of NP;
  (e) P is a determiner of a noun Q and NP is contained in Q.

Sample Output of the Evaluator

(3,6) Missed: its
(3,18) Errors:
(2,12) they
(2,7) they
(2,12) they
(2,2) they
(1,8) Eastern
(3,18) Errors:
(2,12) they