

(Linguistic) Science Through Web Collaboration in the ANAWIKI Project

Udo Kruschwitz, Jon Chamberlain, Massimo Poesio
<http://www.anawiki.org>

1. Problem Overview

Sophisticated *Natural Language Processing* applications are based on massive amounts of (annotated) training data.

Some annotation can be (largely) automated, e.g. part-of-speech tagging.

For more complex linguistic phenomena like **anaphora resolution** there are no tools that result in high-quality annotations without user intervention.

Annotated corpora of the size needed for modern computational linguistics research cannot be created by small groups of hand annotators.

The result is a **resource bottleneck** (i.e. there is not enough training data to make significant improvements).

2. ANAWIKI: Objectives

Demonstrate how the resource bottleneck can be addressed in anaphora resolution tasks.

Strike a balance between collecting annotations from experts (high quality) and from the general Web population (high quantity).

Explore to what extent expert annotations can be substituted by a critical mass of non-expert judgments.

3. Why Web Collaboration?

A huge success story: *Wikipedia* is perhaps the best example of collaborative resource creation.

More recent successful examples: **games with a purpose** such as *ESP game*, *Peekaboom*, *OntoGame* and *Categorilla*.

ESP game example (to collect image labels): 13,500 users played the game, creating 1.3M labels in 3 months.

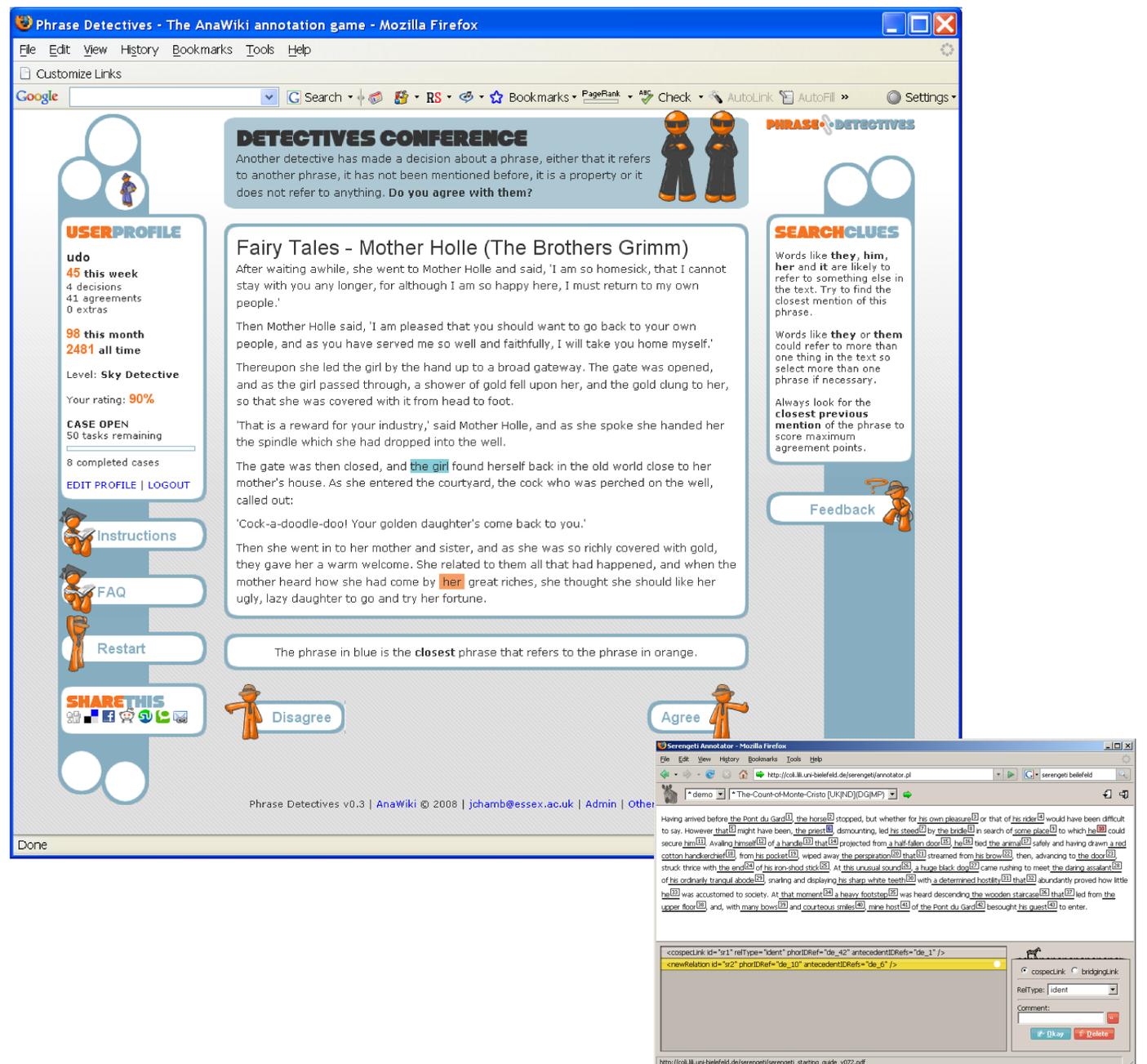
ANAWIKI taps into this huge resource.

4. Experts vs. Non-Experts

Experts: annotate data using **Serengeti** (an expert interface developed by the University of Bielefeld). This provides Gold standard data.

Non-experts: use the **Phrase Detectives** game to annotate texts.

5. Expert and Game Annotation



6. Quality Issues Arising

Do players understand the task?

All players are trained and rated at regular intervals.

Do attention slips affect the quality of collected annotations?

We identify outliers and mistakes are corrected by validation.

What can be done against malicious behaviour?

Players are motivated to enter high quality decisions.

What about genuine ambiguity?

Genuine ambiguity helps us train NLP systems.

7. Does it Work? Yes, It Does!

Full version of Phrase Detectives went live in December 2008.

Over 550 players have contributed more than 130,000 annotations and validations of surprisingly high quality.

Data to be made available via the Anaphoric Bank: anaphoricbank.org

8. Acknowledgements

This research is supported by UK EPSRC, grant EP/F00575X/1.

Thanks also to our colleagues in Bielefeld: Daniela Goecke, Nils Diewald, Maik Stührenberg, Daniel Jettka et al.

ASK US FOR A DEMO!

or try the game at:
phrasedetectives.org