A comparative study of
I2B2 Shared Task 2011
Co-reference resolution in clinical text

UHD University of Houston-Downtown

David Hinote, Carlos Ramirez, Ping Chen

University of Houston-Downtown

Introduction

Co-reference occurs when multiple concepts in a sentence or document refer to the same thing. The purpose of this study is to find an automated process to mark this kind of relation in medical documents, and participate in the I2B2 Shared task in 2011.

Example Co-reference Chains

My boss told me I must give him my final Report.

Approach

Since there are co-reference resolution systems that already exist, we searched for and obtained publicly available tools which mark co-reference. In addition to these tools, we constructed our own rule-based algorithm for marking co-reference which we could specialize for the clinical records. The rest of the study would involve comparing the results of the systems and find which system, or combination of the systems, produces the greatest amount of correct co-referent links.

Building an Algorithm

The first step in building our own algorithm was to construct an environment in which we can easily view the input documents & concepts, and the output co-referent links being discovered by the rules we create. Once finished, we used this tool to write rules for our algorithm, and view the results.

Sample file open in the GUI.
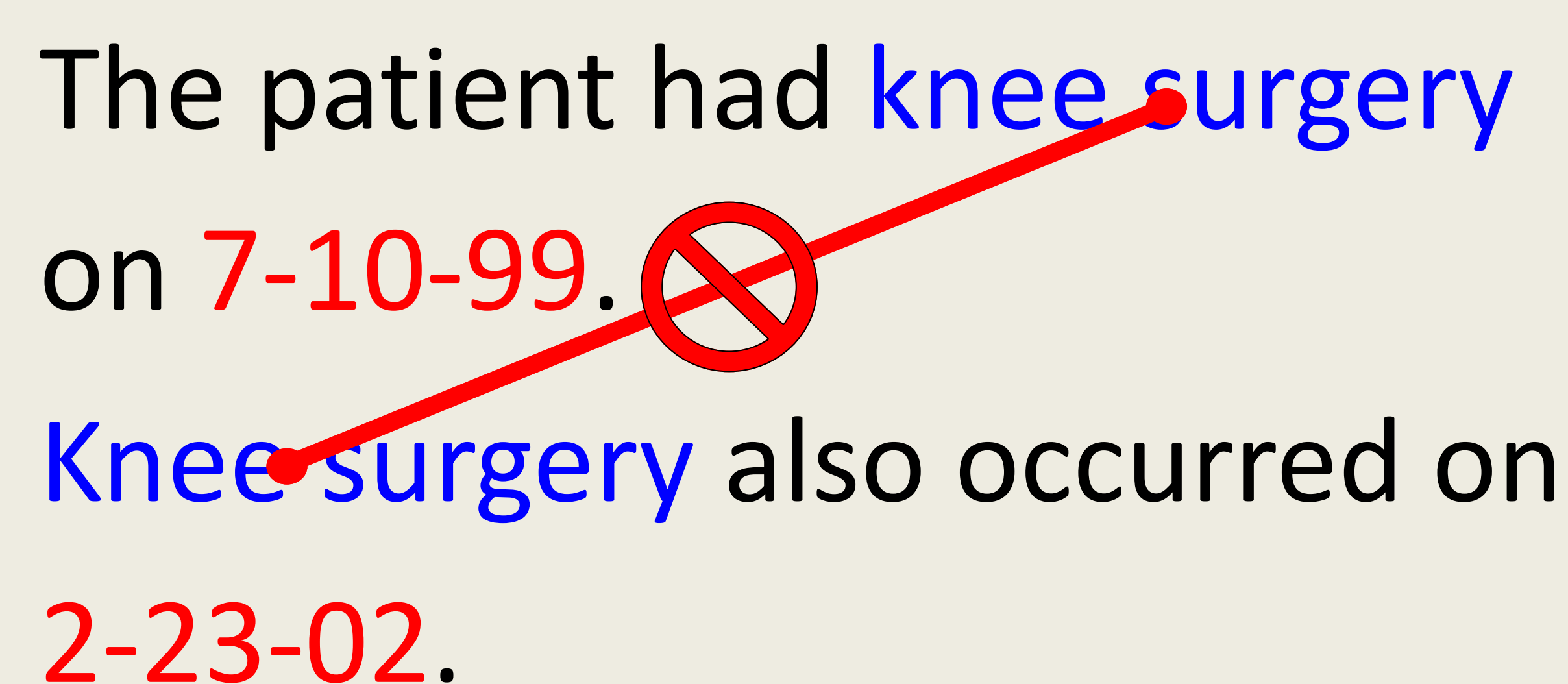
Rules of Co-reference

Using the visual tool, we coded rules that utilize simple string matching, the UMLS, and WordNet databases to give meaning to the concepts and match the meanings. All pronouns use linking rules specifically coded for each case.

String Matching	Syncope ➡ Syncopal Pulmonary embolus ➡ PE
UMLS Matching	Kidney ➡ UMLS Database ➡ C011773 = Renal ➡ ➡ C011773
WordNet Synonyms	Infected ➡ WordNet ➡ 41316 = Septic ➡ ➡ 41316

Link Filtering

After linking concepts that have the same meaning, links between concepts which do not actually refer to the same entity must be filtered out.

The patient had knee surgery on 7-10-99. 🚫

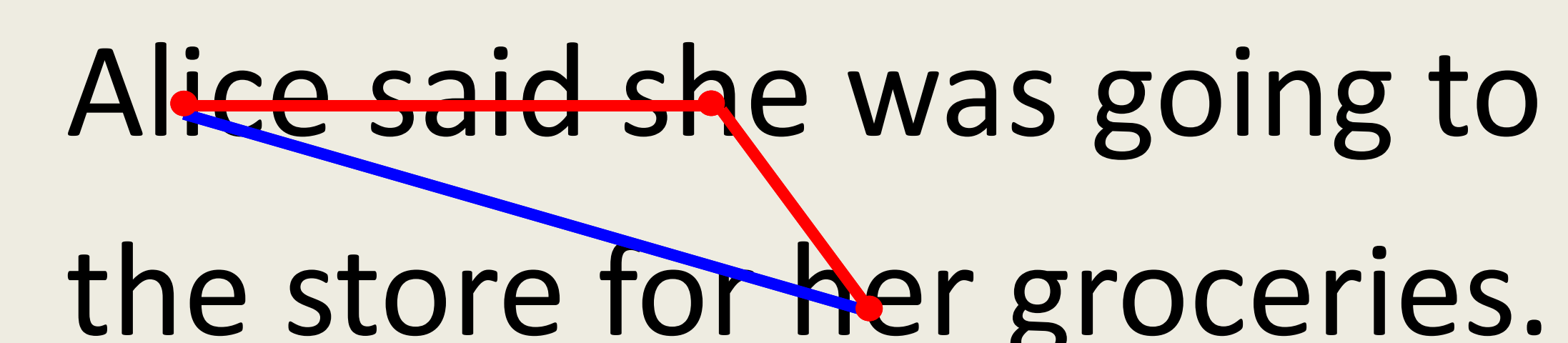Knee surgery also occurred on 2-23-02.

The sentences surrounding the linked concepts are examined for information that indicates if they are actually different entities. If any relevant information is found and it differs, the link is discarded.

Building Chains

Concepts are first linked in pairs, then, after filtering, unnecessary links are removed to make he chains.

Unnecessary Link (in blue)

Alice said she was going to the store for her groceries.

Other Systems & Results

The other systems tested in addition to the rule based system we created are the Stanford NLP System, BART, and LingPipe. The overall scores for all four were obtained using an evaluation script provided by I2B2, the hosts of the 2011 shared task. The following scores are the overall F1 scores of performance on the I2B2 data. The F1 score is the harmonic average of precision (total correct links/total generated links) and recall (total correct links/number of actual links).

UHD	BART	Stanford	LingPipe
0.900	0.775	0.633	0.633

Combining results

The only publicly available co-reference system which found correct links that our algorithm did not was the BART system. We took the output from BART and put it in a union with our output and ran the combination in the script. The result of that was about a 1% increase in recall, but a decrease of about 15% in precision. Because of this, only the results of the algorithm we constructed were sent in for our participation in the 2011 I2B2 Shared Task.

Acknowledgements