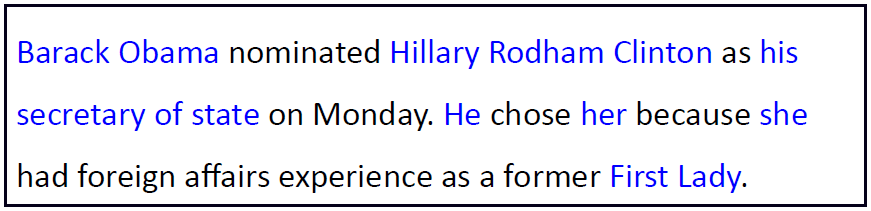
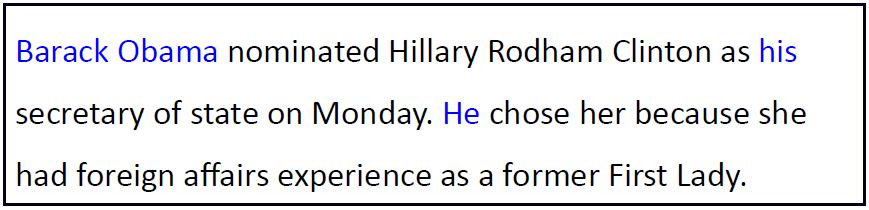
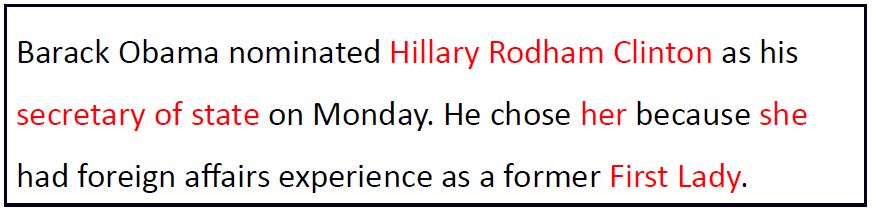
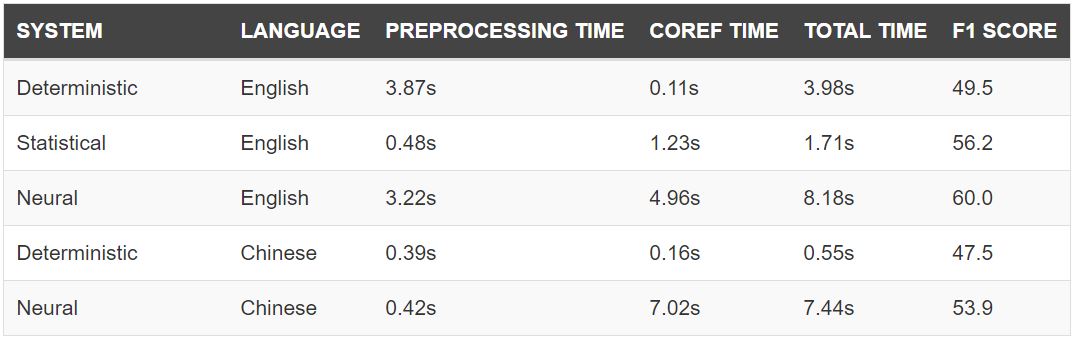
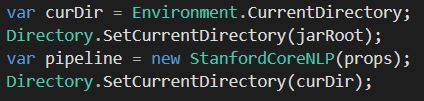
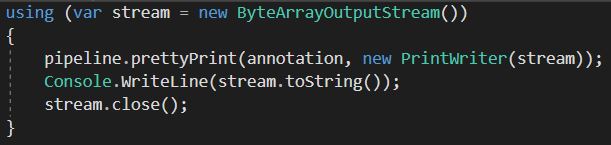
**Coreference resolution**

* Chức năng: Trong những cụm từ nói đến các thực thể của thế giới thực, nhiều cặp từ giống nhau hoặc tham chiếu đến nhau, có một số được lồng vào bên trong thực thể khác, Coreferences resolution xác định tất cả các cụm từ đề cập đến cùng một thực thể. Ví dụ, trong đoạn văn bản:
* Từ “He” và “his” cùng đề cập đến một thực thể là “Barack Obama”
* Từ “her” và “she” cùng đề cập đến thực thể là “Hillary Rodham Clinton”
* Ý nghĩa: Giúp diễn giải văn bản một cách hoàn chỉnh hơn về ngữ nghĩa. Công việc phân loại, tìm kiếm trở nên chính xác hơn.

**Coreference resolution trong Stanford CoreNLP**

* Coreference resolution trong Stanford CoreNLP dùng CorefAnnotator để giải quyết bài toán trên. CorefAnnotator tìm những từ cùng đề cập đến một thực thể trong văn bản. CorefAnnotator thực thi phân giải đồng tham chiếu cả đại từ và danh từ. Toàn bộ từ tham chiếu với những từ được đề cập đầu tiên sẽ được lưu vào CorefChainAnnotation.
* Có 3 hệ thống phân giải tham chiếu khác nhau có sẵn trong CoreNLP
* Deterministic: Phân giải đồng tham chiếu dựa trên quy tắc nhanh (quy tắc sàn lọc nhiều điểm) trong tiếng Anh và tiếng Trung Quốc.
* Statistical: Phân giải đồng tham chiếu dựa trên máy học cho tiếng Anh. Không giống hệ thống khác, nó yêu cầu phân tích sự phụ thuộc cú pháp. Mô hình xếp hạng các mention dựa trên các đặc trưng lớn, nó hoạt động bằng cách lặp qua từng mention trong tài liệu, có thể thêm liên kết giữa mention hiện tại với mention trước đó ở mỗi bước.
* Neural: Phân giải đồng tham chiếu dựa trên mạng nơ-ron, thường chính xác nhưng lại xử lý chậm với tiếng Anh và tiếng Trung Quốc. Mô hình xếp hạng mention dựa trên mạng nơ-ron.
* Ngoài ra còn có hệ thống thứ tư là hybrid or hcoref, nhưng nó không còn được hỗ trợ và cung cấp model trong các phiên bản hiện tại
* Một số chỉ số về hiệu năng khi thử trên Macbook Pro with a 2.4 GHz Intel Core i7 processor
* Cách thực hiện:
* Vào <https://stanfordnlp.github.io/CoreNLP/download.html> download Model jar English (KBP) về máy. Dùng NuGet cài đặt Stanford CoreNLP.
* Tạo đường dẫn đến file StanfordCoreNLP.properties
* Thêm “coref” vào cuối annotators.
* StanfordCoreNLP tự động tìm tất cả các file model
* Thực hiện Annotation lưu trữ phân tích một văn bản. Nó ở dạng map. AnnotationPipeline sẽ đọc lần lượt các annotator, thực hiện một số phép phân tích ngôn ngữ tự nhiên và trả về kết quả. Kết quả trả về được lưu trữ bởi các khóa khác nhau trong Annotation để tích lũy, không ghi đè.
* Đưa ra kết quả tìm kiếm gồm các token được xét theo từng câu của văn bản và đưa ra mối liên hệ giữa các từ cùng đề cập đến một thực thể.

Tham khảo:

<https://stanfordnlp.github.io/CoreNLP/coref.html>

<https://stanfordnlp.github.io/CoreNLP/pipelines.html>

<https://github.com/sergey-tihon/Stanford.NLP.NET/issues/77>