

Abstract

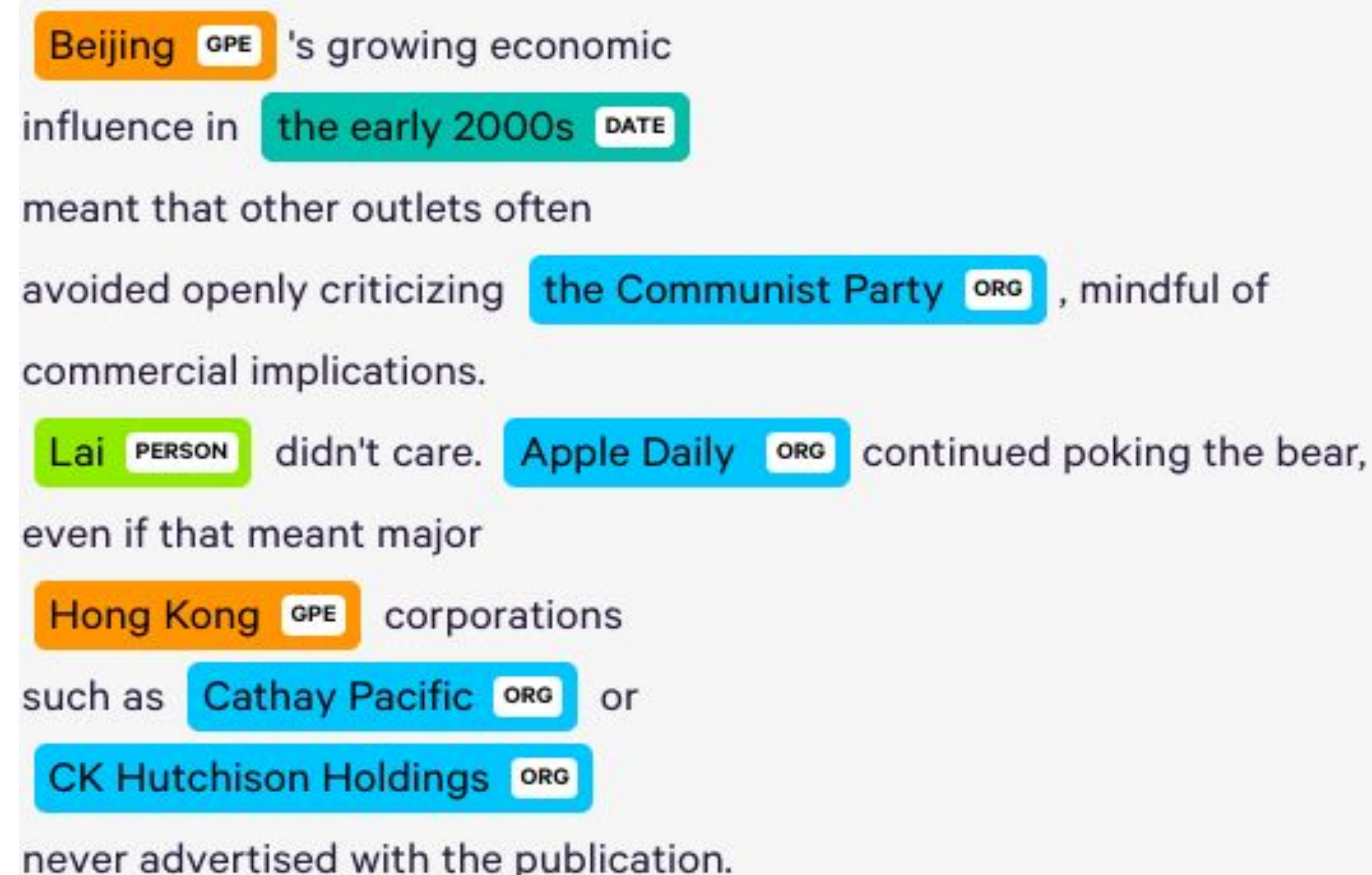
Different news channels have their own styles and perspectives embedded in their coverage of local and global events. The distinctions in news coverage from different channels on a particular event can be analyzed through the unique themes and focus highlighted in their published articles. My project goal is to automate the process of tagging various themes to news articles published by different news sources in order to help news readers get informed on source specialization for specific topics.

Research Goal

To further extend the existing research efforts in news recommendation and fake-news detection systems, we propose a system designed to determine which news sources specialize in covering specific topics and events. Using machine learning and artificial intelligence technology, the proposed system learns to consider several parameters such as coverage frequency and coverage thoroughness when identifying news source specialty.

Methods

We processed over 5000 news articles from roughly 29 different news sources through a Natural Language Processing (NLP) pipeline in Python and assigned a coverage score to each news article. The scores are calculated by identifying if certain news articles included particular key words or phrases. Therefore, the scores help describe the amount of detail provided in the coverage. We used these scores, along with how often a news source covered a particular topic, to make inferences about the topics that certain news sources cover best and prefer to cover.



Beijing **GPE** 's growing economic influence in the early 2000s **DATE** meant that other outlets often avoided openly criticizing the Communist Party **ORG** , mindful of commercial implications. Lai **PERSON** didn't care. Apple Daily **ORG** continued poking the bear, even if that meant major Hong Kong **GPE** corporations such as Cathay Pacific **ORG** or CK Hutchison Holdings **ORG** never advertised with the publication.

Figure 1: Visual representation of named-entity recognition and entity extraction.

Results

We identified which news source published the most coverage and which news source scored the highest for each topic.

Topic	Most Common Source	Highest Scoring Source
Heat wave	independent.co.uk	independent.co.uk
J & k delimitation commission	indiatimes.com	indiatimes.com
Trump organization	nbcnews.com	independent.co.uk
China's communist party	latimes.com	nytimes.com
Australia covid live update	cnbc.com	theguardian.com
Britney spears' request	usatoday.com	foxnews.com

Table 1: Small example of data that outlines sources which most commonly provide coverage and generate the highest coverage scores per topic.

We can use this information to make conclusions about the interests and focuses of a particular news source. For instance, India Times published the most coverage and scored the highest for the Jammu and Kashmir Delimitation Commission topic, which can be expected since Jammu and Kashmir is a region in India. After further analysis of India Times, it can be concluded that India Times specializes in providing coverage for events that occur in India. We applied a similar analysis to the rest of the sources: The Independent and The New York Times are found to specialize in covering politics while CNET specializes in covering matters relating to technology.

Acknowledgments

I would like to thank my mentor Philippe Laban and my PI John Canny for allowing me to participate in their research this summer, as well as Professor Marti Hearst for allowing me to be a part of her lab group. I would also like to thank the E3S and TTE REU staff for this research internship opportunity.

Contact Information

Phone: (510) 401-7755
Email: edreessaied@berkeley.edu

Support Information

This work was funded by the Hopper Dean Foundation.