Understanding Reader Backtracking in Online News Articles

Uzi Smadja, Max Grusky, Yoav Artzi, Mor Naaman

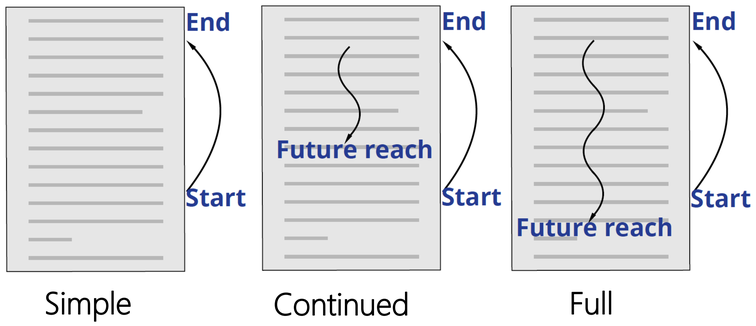
TECHNION Israel Institute of Technology

CORNELL TECH

What is backtracking?

The action of scrolling back in a browser while reading an online news article.

Types of backtracking (BT) events



Why might backtracking occur?

Readers **re-read a piece of text** that they did not fully understand, indicating **readability issues**

Data

Obtained from Chartbeat

1.4M sessions

26k users

8k online news articles

two major online publications:
long-form news magazine (LONG)
short-form news website (SHRT)

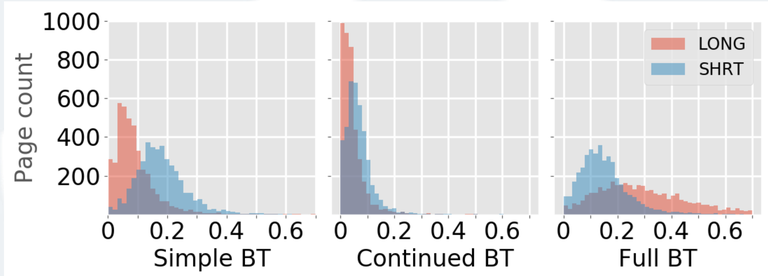
Are some pages more prone to backtracking?

Full backtrack events are more common, especially on long-form pages.

Number of backtracking events for long and short articles:

	Users	Pages	Sessions	Simple BT	Cont. BT	Full BT
LONG	15,949	3,968	694,728	57,897	28,840	246,787
SHRT	11,117	3,957	763,335	136,153	49,211	127,421

Mean number of backtracking events per reading session:



What textual features predict backtracking?

Lexical: readability measures & simple text properties

Entity-density: proportions of named entities

Part-of-speech density: proportions of nouns

Coreference chains: characteristics of coreference chains

	Accuracy	F1	AUC
All features	**0.838**	**0.839**	**0.901**
Length only (control)	**0.747**	**0.739**	**0.809**

Find in the full paper:

- Feature ablation showing relative contribution of each family of features
- How we control for different types of readers and topics, finding no effect on the prediction task

What does this mean for Web content and readability?

- Full backtrack events may be an indication of readability issues in online news.
- Large-scale Web signals can help define improved readability measures.

Learn more:

bit.ly/backtracking-www

Contact: uzi.smadja@gmail.com

Support: Yahoo! Research, NSF IIS-1840751