



## Where Big Data Meets Real-Time: Efficiently Indexing and Ranking News using Activity

**Speaker:** Paul Ogilvie

**Date:** Thursday, Feb 9, 2012

**Time:** 3:30-4:30pm

**Location:** GHC 6115



### Abstract:

Having access to large quantities of data can transform the way you approach a problem. The recent fervor in industry around big data is not without justification. For example, knowing who is sharing or reading a news article can be as valuable as the content of the article itself. Consider identifying timely news relevant to an industry, such as the Internet industry. Simple counts of LinkedIn members in the Internet industry who have shared, tweeted, or read an article are important signals of relevance.

Many solutions to handling large scale data focus on periodic batch processing using distributed Map-Reduce techniques. News is ephemeral, so the delay introduced by batch processing solutions may sacrifice much of an article's shelf life. This talk presents a novel use of inverted lists, a common data structure used by traditional search-engines, to efficiently index and support real-time updates to document (article) representations.

### Speaker's Bio:

Paul Ogilvie is the Engineering Lead for news relevance at LinkedIn. This team is responsible for the infrastructure and models used in LinkedIn Today, LinkedIn's news aggregation product. Prior to LinkedIn, Paul worked at a news personalization start-up in Pittsburgh, mSpoke, which was acquired by LinkedIn in 2010. Paul has a Ph.D. from the Language Technologies Institute at Carnegie Mellon University, where he studied the use of document structure and annotations on a diverse range of search tasks.

[pogilvie@linkedin.com](mailto:pogilvie@linkedin.com)