wikiFeed: Keeping Wiki Content Current via News Sources
Rachel Adams, Alex Kuntz, Morgan Marks, William Martin, David R. Musicant
Carleton College, Northfield, MN, USA

Overview
- Maintaining current and reliable information in a wiki can be challenging
- wikiFeed assists wiki contributors with this task
- Helps users track news, makes recommendations as to where content should be integrated into wiki, and assists users in citing sourced content
- Demonstration built on English Wikipedia

Workflow and Interface Design
- Design goal: Help wiki contributors identify new content, and make recommendations to make it easy to add that content as well as cite source content appropriately
- News content sources: RSS and Twitter feeds
- Headlines and content pulled down on initial setup and every 30 minutes
- User interface resembles RSS reader
- Headlines presented, and are clickable to see source news article
- On request, a list of 20 automatically recommended Wikipedia article titles for which news content may belong is shown
- Sorted in descending order by their relevance to the news source
- Wikipedia article titles are clickable, and open up the associated Wikipedia article
- Below the list of relevant Wikipedia articles, wikiFeed presents an automatically generated citation to that news source appropriately formatted in Wiki markup

For a Wikipedia editor looking to add content, this setup provides a streamlined and direct way to assist in finding and adding new information.

Implementation
- Implemented in Python using Django web framework, running on standard Apache web server
- RSS and Twitter feeds contain headlines and links, which are extracted according to format
- Actual text of a news source obtained by extracting whatever text is contained in the first link in headline
- Sphinx (http://sphinxsearch.com/) used for efficient matching
- Static recently available current-text Wikipedia dump used instead of online Wikipedia, for efficiency; content may change, but matching is more stable over time
- Keyword-based matching algorithm to find the most closely related Wikipedia articles to the news source of interest
  Specifically, two-step process for article matching:
  1. Find important keywords in news source: remove stopwords, then select the ten words from the news source with the highest tfidf values.
  2. Use BM25 ranking function to rank similarity between a Wikipedia article and a the news source. Two different BM25 values are determined: one for the title of the Wikipedia article, and one for its body; weighted average is taken.
Based on above techniques, we construct a list of the 20 Wikipedia articles that best match a given list of keywords.