

Visual Interactive Exploration of Online Discourses with TopicExplorer

Alexander Hinneburg¹, Christian Oberländer²

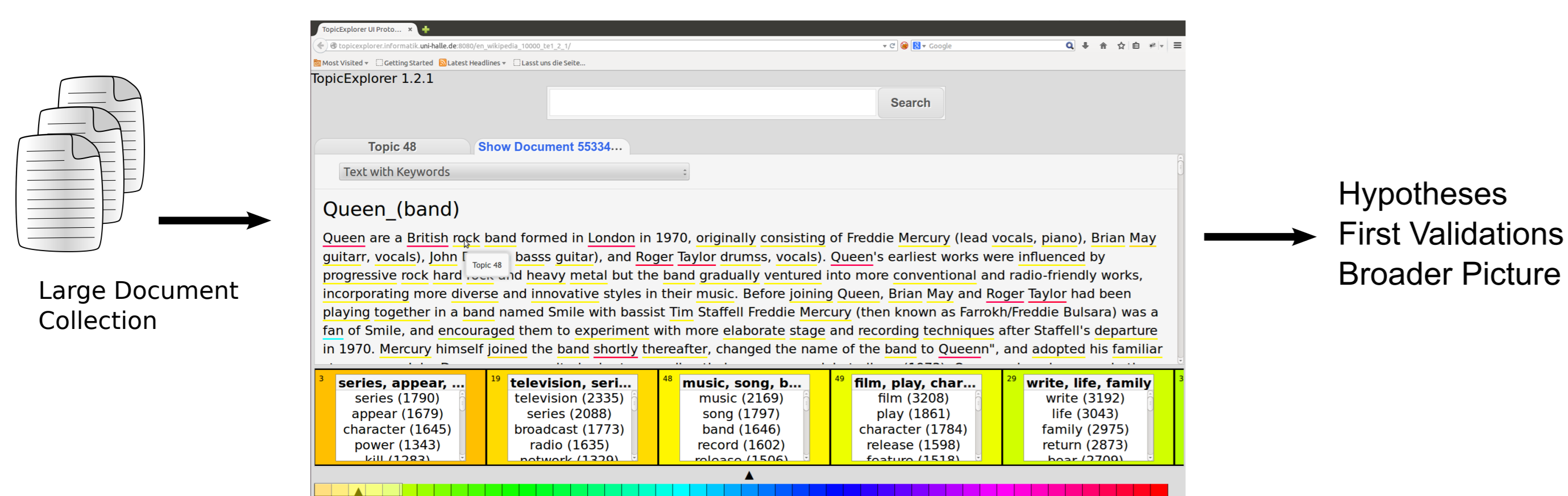
¹Computer Science, ²Political Sciences and Japanese Studies, Martin-Luther-University Halle-Wittenberg

<http://topicexplorer.informatik.uni-halle.de/>

Problem Statement

- Large document collections contain interesting knowledge and stories
- Humans are too slow to read all documents
- Machine learning and topic models can help to explore large document collections pursuing similar goals similar to Grounded Theory
- Challenges
 - Topic models compute a lot of potentially useful information. How to turn topic models into an efficient tool for exploration?
 - Topic models give no guarantees that topics are interpretable. How to go from statistical models to semantic interpretation?

Application Context of TopicExplorer

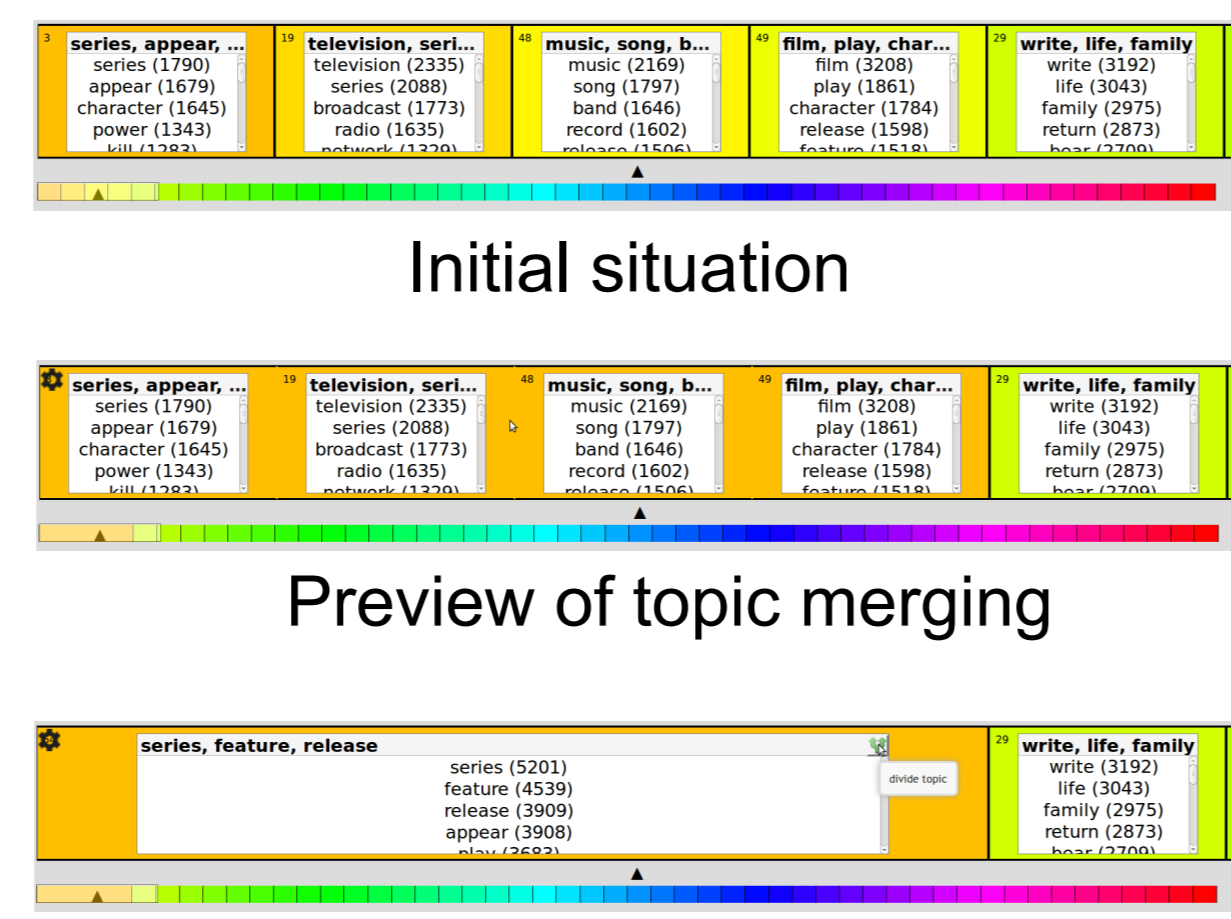
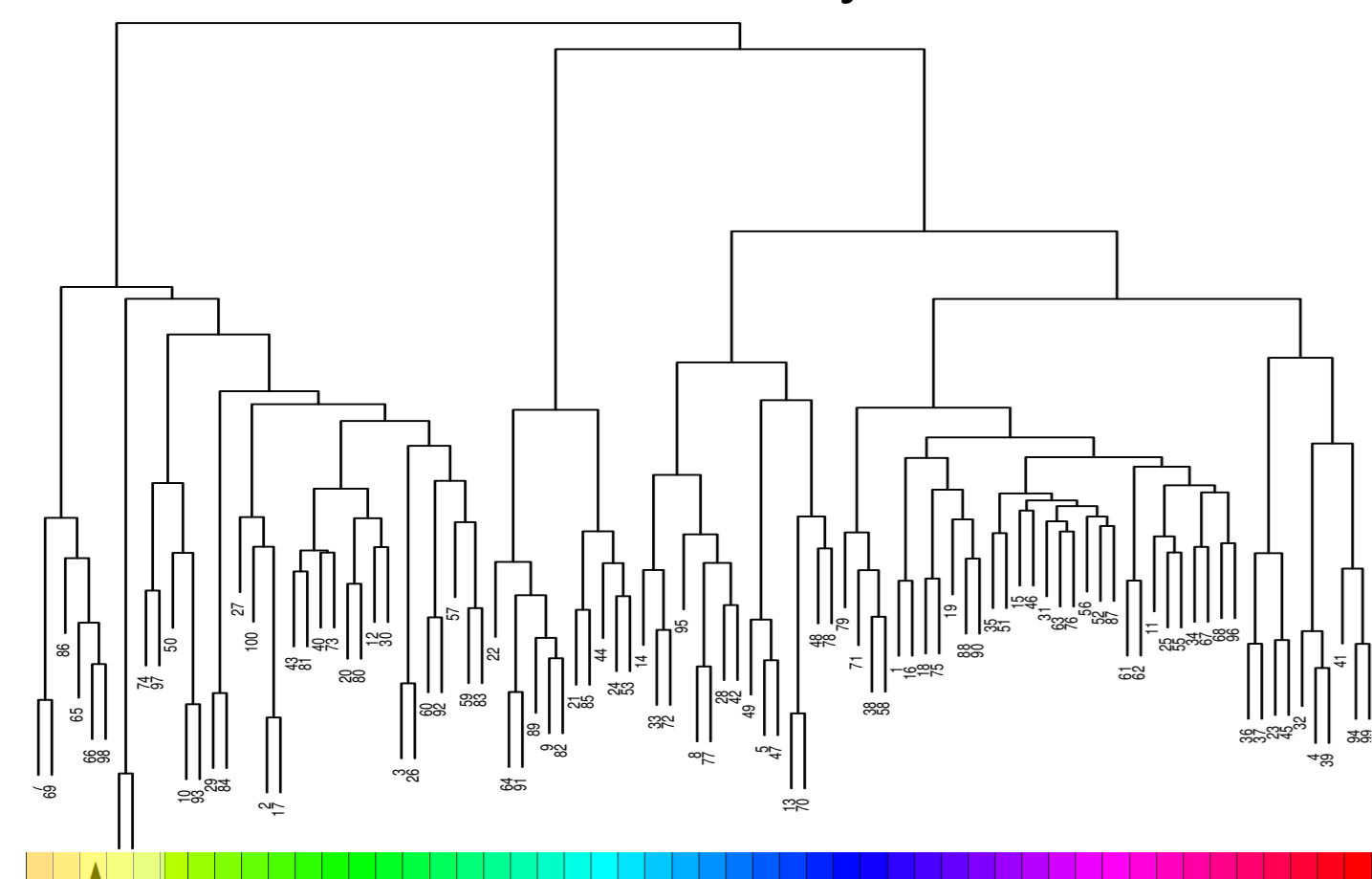


Key Features

Linear Topic Ordering, hierarchical Topics

Hierarchical Clustering of Topics based on Word Distribution Similarity

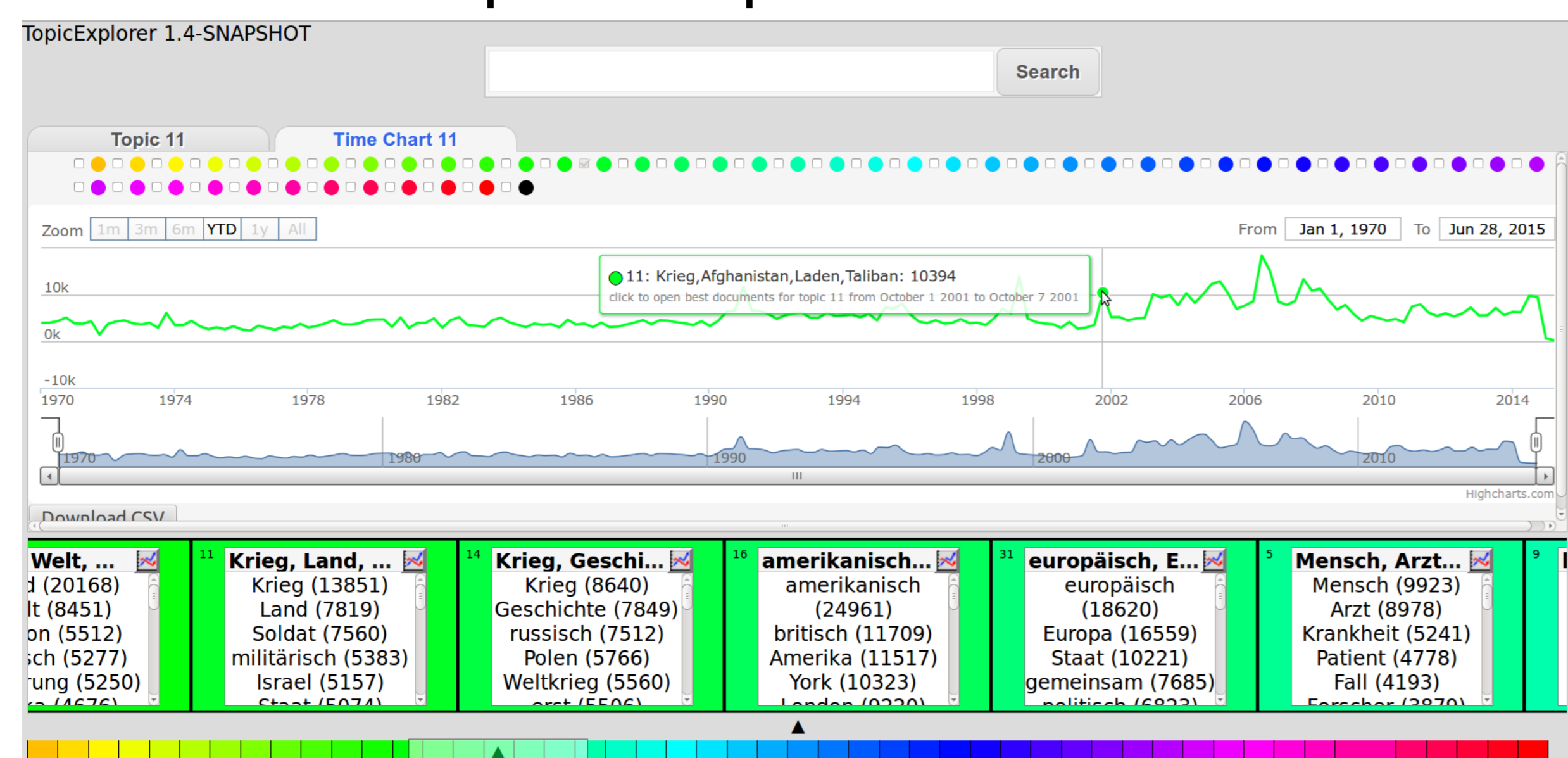
Navigating Hierarchical Topics



Linear Ordering of Topics => Color Mapping

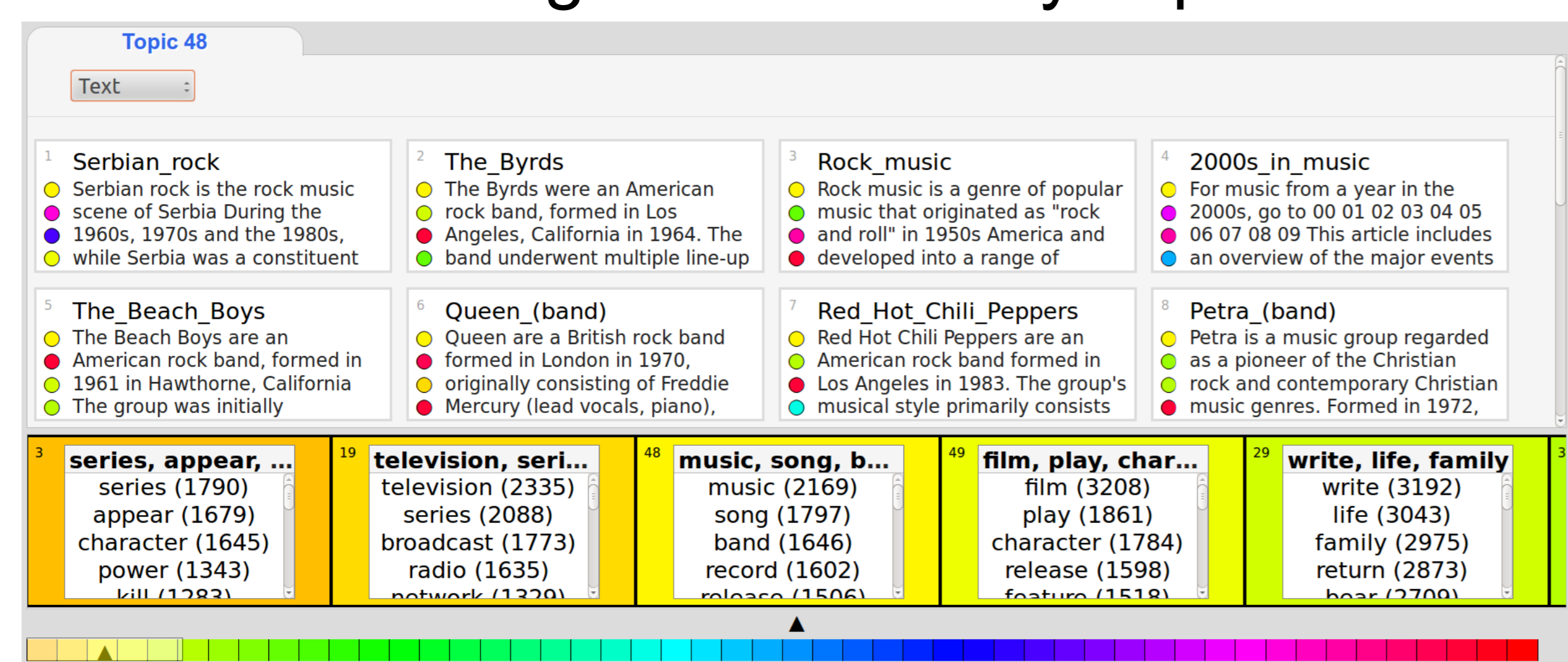
- Linear order enables semantically smooth topic browsing
- Color acts as a visual hash function that helps recognize topics in different contexts
- Hierarchical topics allow to choose topic granularity

Temporal Topic Evolution



Example: a German newspaper corpus

Ranking Documents by Topic

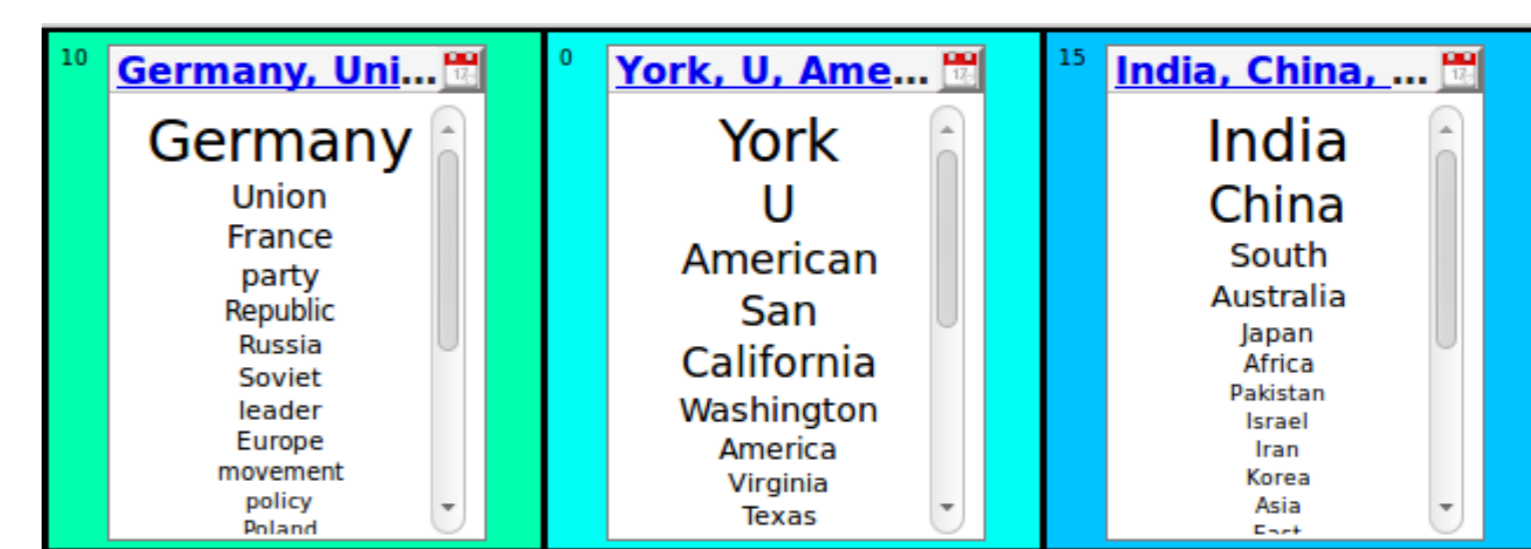


Example: subset of English Wikipedia, Ranking by Topic "music, song,..."

Topic Frames

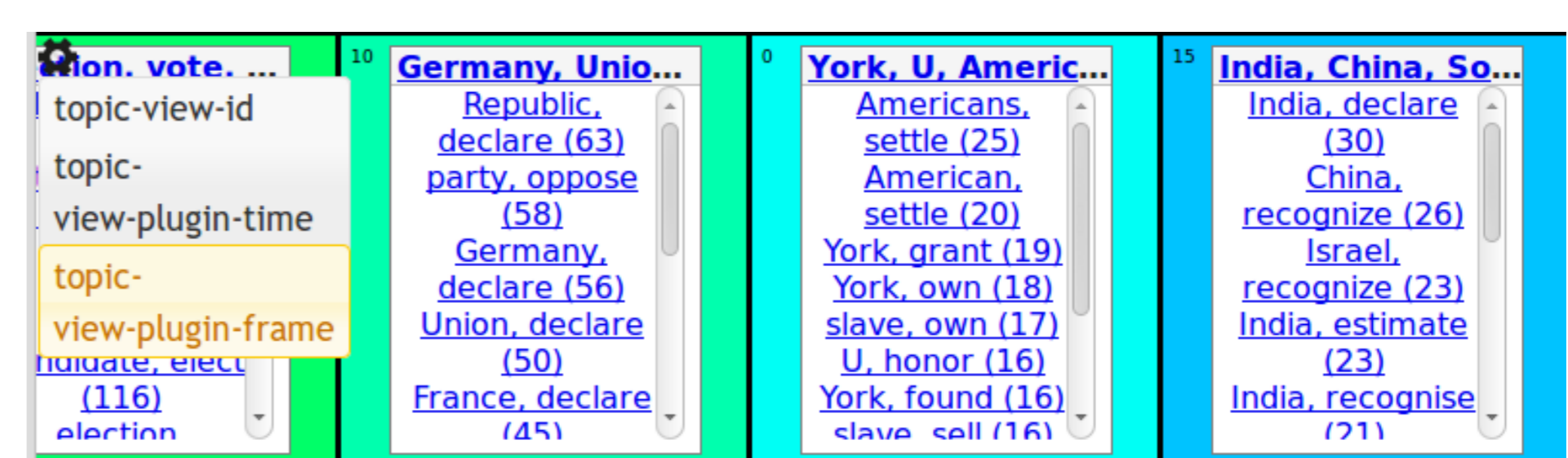
A **Topic frame** occurs when a noun and a verb appear close together in a document and both tokens are assigned to the same topic.

Ambiguous Topics



Topics are focused on some geographic region. However, they can be ambiguously interpreted as tourism, politics or history.

Topic Frames



Topic frames allow more clear interpretation of topics, e.g. the India-China topic is focused on foreign politics, whereas the America topic is focused on history.

Topic Frame Occurrence

lect furs and control the upper Columbia River fur trade. Fort Nisqually was rd. Fort Victoria was erected in 1843 and became the headquarters of of British Columbia. By 1840 the HBC had three forts: Fort Hall (purchased the Oregon Trail route as well as Fort Vancouver near its terminus in the eded aid to the early Oregon Trail pioneers. When the fur trade slowed in ritish was seriously diminished. Canada had few potential settlers who were British and American, and their families did start settling in Oregon, nada. In 1841 James Sinclair on orders from Sir George Simpson guided yer and Red River near present Winnipeg, Canada into the Oregon territory. ette Valley, with their promise of free land and HBC-free government. In British lost the land north of the Columbia River they had so long controlled.

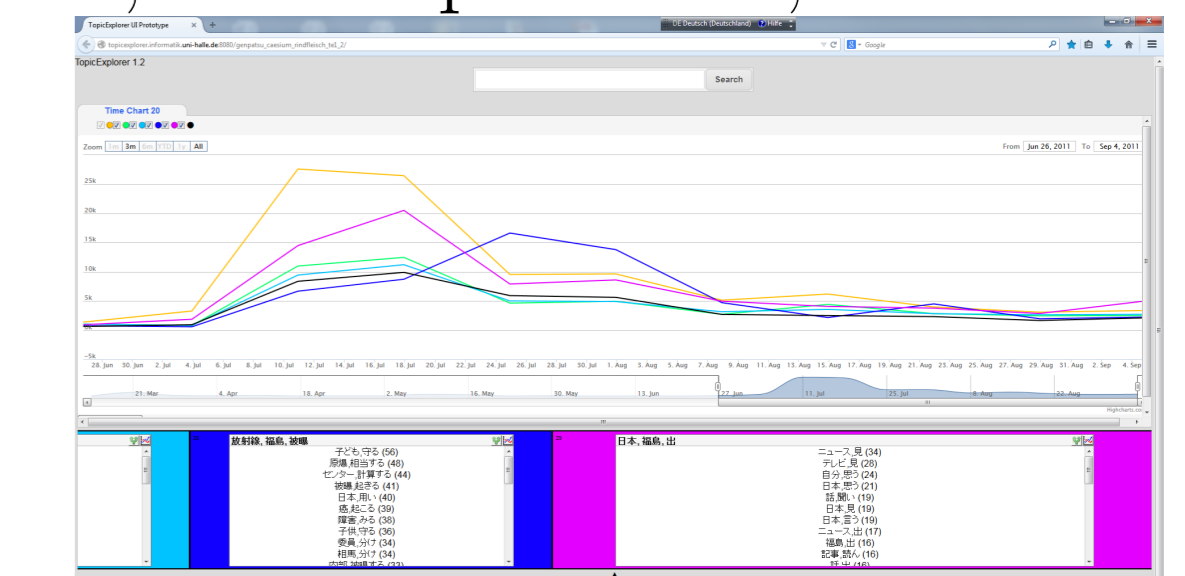
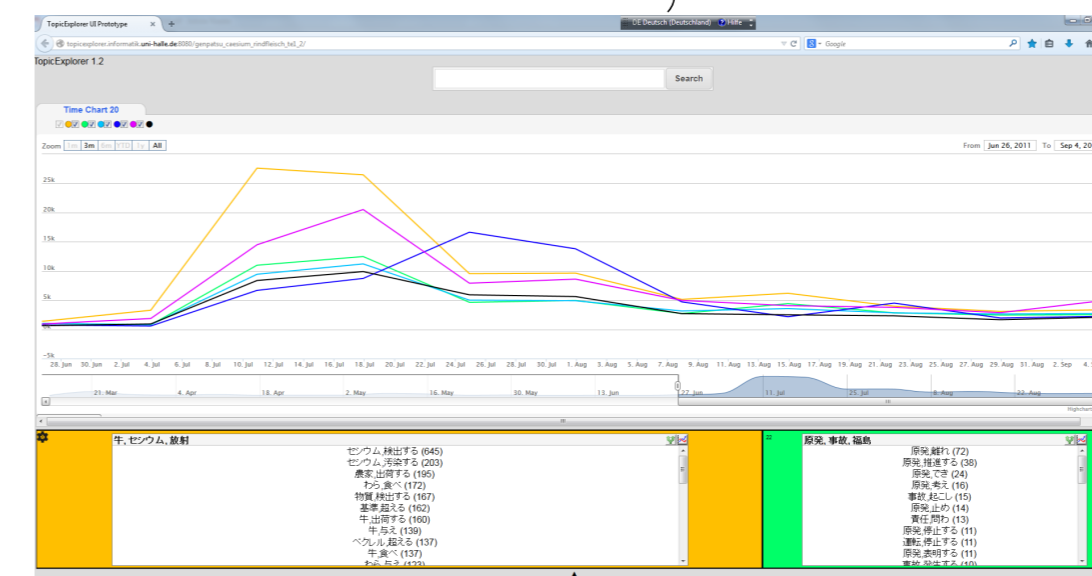
Application Use Case: Japanese Blogs

- Online discourse about Cesium-contaminated beef scandal after the Fukushima nuclear accident of 2011
- Blog posts from blog aggregator (Kizashi) containing the words nuclear (Genpatsu) and beef (Ushi)
- Related press analysis available
 - Toru (2012): An Exploratory Study of Media Frames in Quantitative Text Analysis. Through the Newspaper Report on the Problems with Radioactive Cesium-contaminated Beef (Japanese). In: Social Information Studies 1 (2), pp. 61-76.
- Course of events
 - 18 March – limits for radioactivity in food announced; controls start
 - 9 July – radioactive Cesium found above limit in beef from Fukushima
 - 19 July – beef sale banned in Fukushima; later extended to Miyagi, Iwate, Tochigi
 - 19 August – sales ban lifted in Miyagi
 - 26 August – sales ban lifted in all prefect.

Three specific Topic-Peaks

Cesium, ...

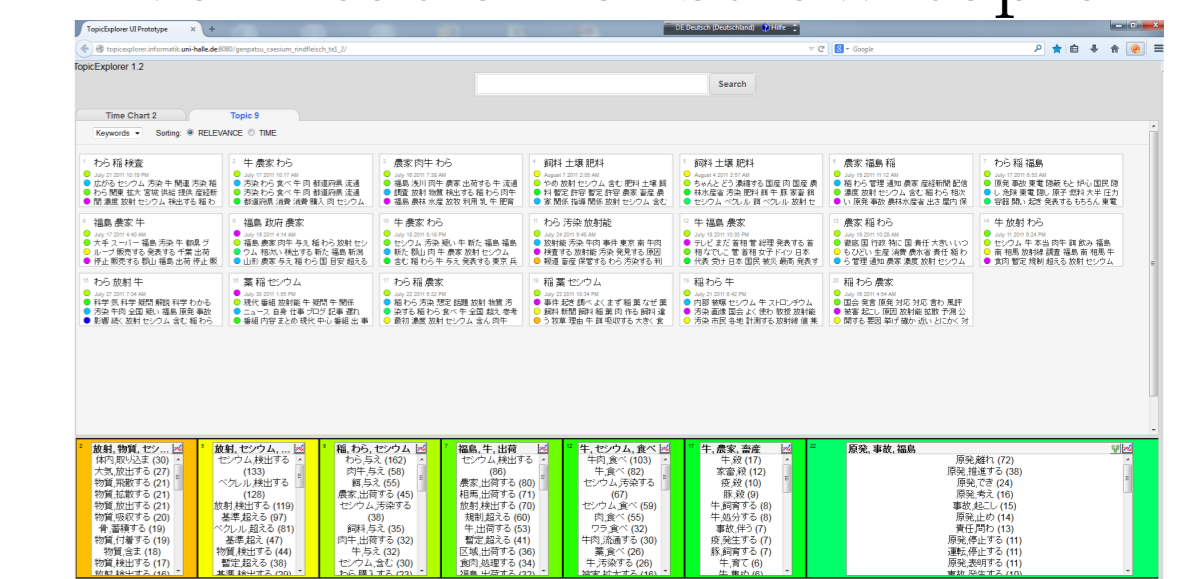
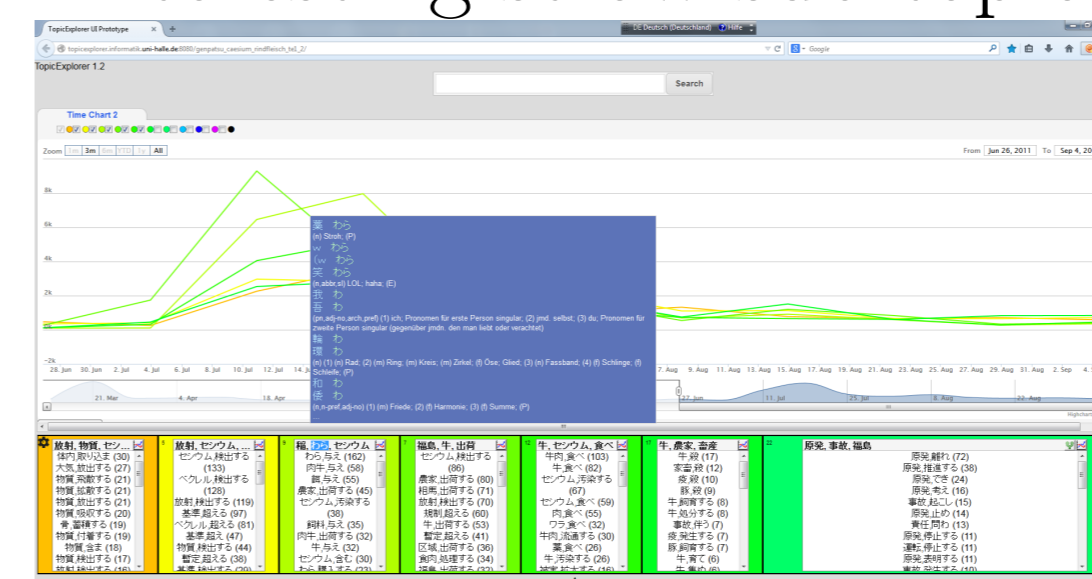
..., health problems, and news



Cesium Topic

interesting straw sub-topic

verification of straw topic



Christian Oberländer: *Digitale Japanologie und neue Medien in der Risikogesellschaft: Der Skandal um radioaktiv verseuchtes Rindfleisch als Fallbeispiel*, to appear in David Chiavacci / Iris Wiczorek (eds.): Japan 2015 Politik, Wirtschaft und Gesellschaft, Berlin: VSJF, 2015, 30 S.