1 Introduction
Organisations and individuals that need to monitor what the media say about certain issues face an extreme information overload, especially if they are interested in the news written in more than one language. News aggregators sometimes pre-filter potentially user-relevant articles or automatically group related articles into clusters. However, the enormous amount of available online information calls for further automatic information processing to enable users to sieve through even larger amounts of textual data in less time and to navigate and explore the document collections efficiently.

2 NewsExplorer
NewsExplorer is a freely available news analysis system that offers such functionality in 19 languages. NewsExplorer integrates various text analysis applications including clustering, multi-label document classification, named entity recognition, name variant matching across languages and writing systems, topic detection and tracking, and more. The purpose of this presentation is to present this news exploration and analysis system and to especially address the multilinguality issue and the cross-lingual functionality of the application. References to prior art will be made, where appropriate.

3 News Data and the EMM family of applications
NewsExplorer is part of the Europe Media Monitor (EMM) family of applications (http://press.jrc.it/overview.html). EMM gathers a daily average of 80,000 news articles from about 2,200 web news sources in 43 languages. NewsBrief and the Medical Information System MedISys classify the news, cluster related articles and alert users of breaking news when unexpected spikes are detected. EMM-Labs gives access to data visualisation tools and to the results of a collection of advanced text processing tools such as relation extraction, event scenario template filling, and various types of social networks. The freely available EMM online applications attract between one and two Million hits per day.

References
Linking News Content Across Languages


Acknowledgements

NewsExplorer and the other EMM media monitoring applications are complex applications, which were built in a group effort. I would thus like to thank the whole EMM team for their good work and especially Bruno Pouliquen, who was the main driving force behind the development of NewsExplorer.

Figure 1. Screenshot of NewsExplorer, showing a map with the location of today’s news, the largest English language news clusters, links to related news in the other 18 languages, lists of countries, people and organizations mentioned in the news that day, and lists of the biggest ‘stories’ (daily news clusters linked over time) this week, month and year.

Figure 2. Screenshot of (part of) NewsExplorer’s page on Pakistani President Pervez Musharraf. That page shows automatically collected name variants (including variants in different scripts such as Arabic, Farsi, and Russian), titles, lists of related persons, lists of related news clusters, quotations by and about Musharraf, multi-day ‘stories’ in which he is mentioned, and more. Relations between two or more persons can be visualised graphically.