The Typed Index

If you want to search in a multilingual environment with high-quality language-specific word-normalization, if search should be case-sensitive, at least as far as scoring is concerned (e.g. you want to distinguish between search for the company "MAN" and the word "man"), if you want to add phonetic search for names and go towards a semantic search which distinguishes between a search for the color "brown" and a person with the second name "brown", in all these use cases you have to deal with different types of terms. I will show why it makes much more sense to attach types to Lucene terms instead of relying on different fields or even different indexes for different kinds of terms. Furthermore I will show how queries to such a typed index look like and why e.g. SpanQueries are needed to correctly treat compound words and phrases or realize phonetic search.

About the Speaker:
As Director Research at IntraFind Software AG, Christoph Goller is responsible for coordinating research activities with research centers, universities as well as within the company. Furthermore, he is responsible for the development of IntraFind´s text analytics components. Christoph got a PhD in computer science from the Technical University of Munich where he was working on artificial intelligence, machine learning and knowledge representation. His main interests are scalable information retrieval, search-based applications, and text analytics. Christoph is Apache Lucene committer since 2004. He has accompanied dozens of commercial projects using Lucene and Solr and has more than 15 years of experience in information retrieval. Christoph is author of more than 15 scientific papers and frequently gives presentations on search related topics e.g. at Lucene Revolution, Berlin Buzzwords, ApacheCon.

http://vimeo.com/26595316
http://vimeo.com/44023462

About IntraFind:
IntraFind Software AG, founded in 2000 supports its over 850 customers with products, solutions and consulting centered on the topics of enterprise search, text analytics and big data search. The technological portfolio includes user-friendly high-performance scalable full-text search, enhanced with linguistic, statistical and semantic modules. Based on these modules IntraFind delivers more relevant search results, semantic context-sensitive search, navigation, and faceting. IntraFind's products for information access to structured and unstructured data are based on the well-known open-source products Lucene, Solr and Elastic Search. Two developers of IntraFind are Lucene committers. As an independent software vendor, IntraFind cooperates with 25 embedding and reselling partners in Europe. Renowned customers are: Audi AG, Robert Bosch GmbH, Deutsche Post AG, Krauss-Maffei Wegmann, Landesbank Berlin, Merck, Rohde & Schwarz, Voith GmbH, Webasto, Wolters Kluwer, Zeit Online, ZF Friedrichshafen.