



The Typed Index

FOSDEM 2015, Open Source Search Dev Room

Christoph Goller

1. About IntraFind and Me (very brief)
2. Basics about Inverted File Index, Lucene Analysis Chain
3. The Typed Index: Different Types of Terms
4. Multilingual Search / Mixed Language Documents / Language Chunking
5. The Typed Index and Semantic Search

Specialist for Information Retrieval and Enterprise Search

- ▶ Founding of the company: October 2000
- ▶ More than 850 customers mainly in Germany, Austria, and Switzerland
- ▶ Employees: 30
- ▶ **Lucene Committers:** B. Messer, C. Goller

- ▶ Independent Software Vendor, entirely self-financed
- ▶ **Products are a combination of Open Source Components and in-house Development**
- ▶ Support (up to 7x24), Services, Training,
- ▶ Focus on Quality / Content Analytics / **SOA Architecture**
 - ▶ Linguistic Analyzers for most European Languages
 - ▶ Semantic Search
 - ▶ Named Entity Recognition
 - ▶ Text Classification

IntraFind Software AG – Customers

INTRAFIND



Audi



ZEIT  ONLINE

VOITH



Basics: Inverted File Index and Lucene Analysis Chain

Inverted File Index

1	The old night keeper keeps the keep in the town
2	In the big old house in the big old gown.
3	The house in the town had the big old keep
4	Where the old night keeper never did sleep.
5	The night keeper keeps the keep in the night
6	And keeps in the dark and sleeps in the light.

Table with 6 documents

<< index >>

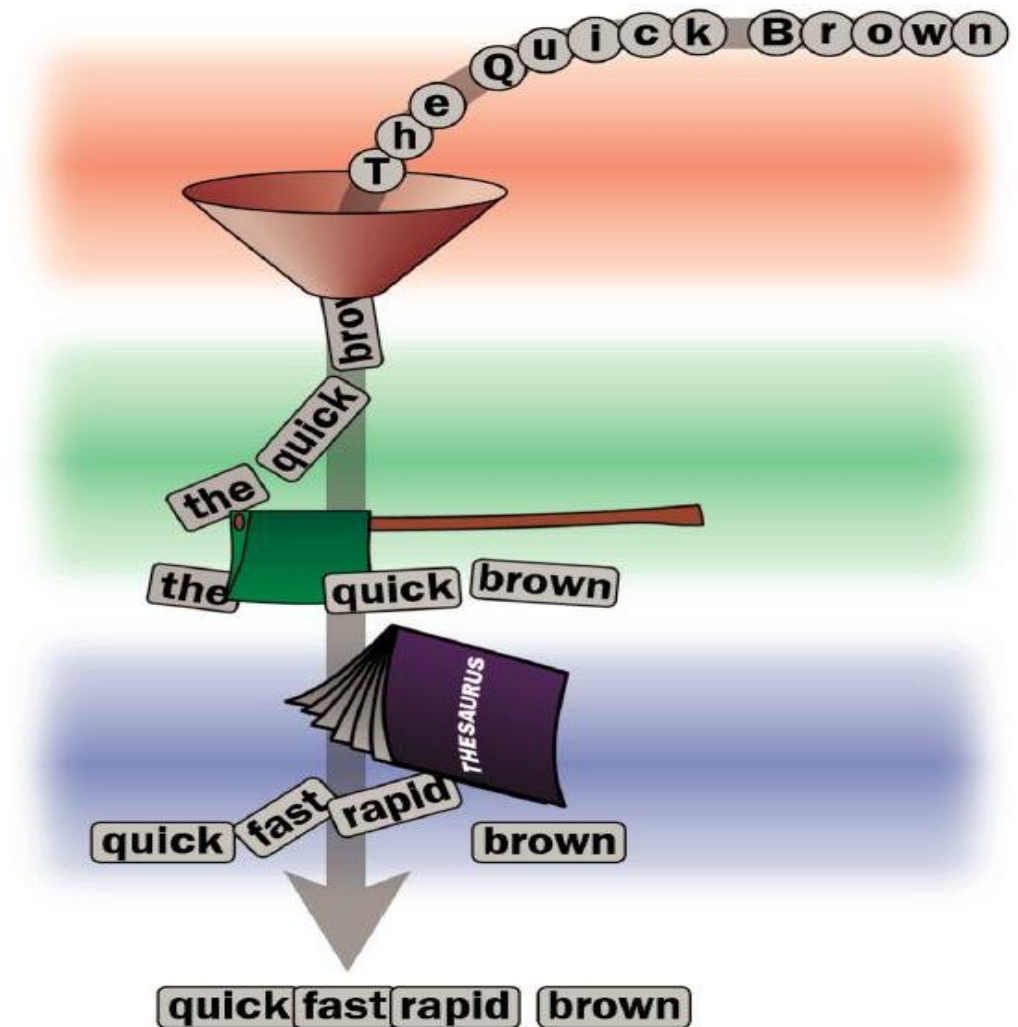
?

Lucene

Term	Freq	Postings & Positions
and	1	(6,1) (6,6)
big	2	(2,3) (2,8) (3,8)
dark	1	(6,5)
did	1	(4,7)
gown	1	(2,10)
had	1	(3,6)
house	2	(2,5) (3,2)
in	5	<(1,8)> <(2,1)> <(2,6)> <(3,3)> <(5,7)> <(6,3)> <(6,8)>
keep	3	(1,7) (3,10) (5,6)
keeper	3	(1,4) (4,5) (5,3)
keeps	3	(1,5) (5,4) (6,2)
light	1	(6,10)
never	1	(4,6)
night	3	(1,3) (4,4) (5,2) (5,9)
old	4	(1,2) (2,4) (2,9) (3,9) (4,3)
sleep	1	(4,8)
sleeps	1	(6,7)
the	6	<(1,1)> <(1,6)> <(2,2)> <(2,7)> <(3,1)> <(3,4)> <(3,7)> <(4,2)> <(5,1)> <(5,5)> <(5,8)> <(6,4)> <(6,9)>
town	2	(1,10) (3,5)
where	1	(4,1)

Break stream of characters into tokens /terms

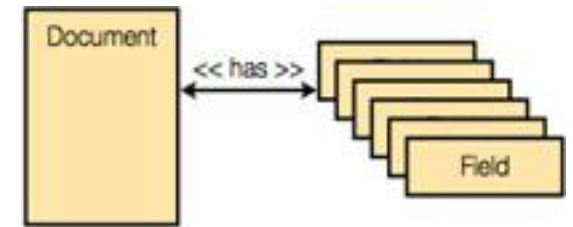
- ▶ Normalization (e.g. case)
- ▶ Stop Words
- ▶ Stemming
- ▶ Lemmatizer / Decomposer
- ▶ Part of Speech Tagger
- ▶ Information Extraction



More about Documents, Fields, and Tokenstream

INTRAFIND

- ▶ A Document consist of Fields
- ▶ Analyzer a TokenStream Factory
- ▶ TokenStream follows a Decorator Pattern
 - ▶ Tokenizer splits text into tokens
 - ▶ TokenFilter processes tokens - operates on Tokens
- ▶ Each TokenStream can register with Attributes of interest
- ▶ Lucene provides core Attributes
 - ▶ TermAttribute - holds term data
 - ▶ PositionIncrementAttribute - holds position information
 - ▶ PositionLengthAttribute – if a token spreads over several positions
 - ▶ OffsetAttribute - holds char offsets
 - ▶ PayloadAttribute - holds arbitrary per term data



Position and Position Length

Term	IBM	International	Business	Machines
Position increment	1	0	1	1
Position length	3	1	1	1

- ▶ Tokens on the same Position are regarded as synonyms
- ▶ **But:** Tokens may spread over several Positions -> Position Length
- ▶ PositionLength are not yet supported in the index

What about word decomposition?

iphone6 -> iphone 6

mother-in-law -> mother in law

German Examples:

Bundesfinanzministerium -> Bundes|finanz|ministerium

Ministerium des Bundes für Finanzen ---- Bundesministerium für Finanzen

Wachstube -> Wach|stube or Wachs|tube

The Typed Index: Different Types of Terms

- ▶ Lemmatizer: maps words to their base forms

English			German		
going	→	go (Verb)	lief	→	laufen (Verb)
bought	→	buy (Verb)	rannte	→	rennen (Verb)
bags	→	bag (Noun)	Bücher	→	Buch (Noun)
bacteria	→	bacterium (Noun)	Taschen	→	Tasche (Noun)

- ▶ Decomposer: decomposes words into their compounds
 - ▶ Kinderbuch (children's book) → Kind (Noun) | Buch (Noun)
 - ▶ Versicherungsvertrag (insurance contract) → Versicherung (Noun) | Vertrag (Noun)
- ▶ Stemmer: usually simple algorithm
 - ▶ going → go decoder, decoding, decodes → decod
 - ▶ Overstemming: Messer → mess king → k
 - ▶ Understemming: spoke → speak

Bad Precision with Algorithmic Stemmer

[\[WEB\] Apache Lucene/Solr - Who We Are](#)

Lucene/Solr is maintained by a team of volunteer developers. Core Committers Bill Au (billa@...) Michael Busch (buschmi@...) Doron Cohen (doronc@...) Doug Cutting (cutting@...) Shai Erera (shaie@...) Erick Erickson (erick@...) Otis Gospodnetic (otis@...) Martijn van Groningen (mvg@...) Erik Hatcher (ehatcher@...) Chris Hostetter (hossman@...) Jan Høydahl (janhoy@...) Grant Ingersoll (gsingers@...) Mike McCandless (mikemccand@...) Ryan McKinley (ryan@...) Chris Male (chrism@...) Bernhard Messer (bmesser@...) Mark Miller (markmiller@...) Robert Muir (rmuir@...) Stanislaw Osinski (stanislaw@...) Noble Paul (noble@...) Steven Rowe (sarowe@...) Uwe Schindler (uschindler@...) Shalin Shekhar Mangar (shalin@...) Yonik Seeley (yonik@...) Koji Sekiguchi (koji@...) Dawid Weiss (dweiss@...) Andi Vajda (vajda@...) Simon Willnauer (simonw@...) Emeritus Committers Josh Bloch Peter Carlson (carlson)

<http://lucene.apache.org/java/docs/whoweare.html>

[\[LUCID\] Information creation proliferation? Forbes looks to Solr and Lucid Imagination](#)

2011-01-12 11:42

Information: we all love to make more of it, but it sure piles up. Forbes Magazine Online blogger Quentin Hardy takes on making sense of it in a nice post about Solr/Lucene and Lucid Imagination. Some soundbites: Our civilization may pride itself on the amount of information we create – more than every conversation, ever, in a couple of years; enough to jack The Library of Congress to the Moon on data-packed CD-Roms, take your superlative – but we've also made a holy mess of it. It is mostly, as they say, "unstructured," meaning as random as your last 25 emails, your tweets, and all those spreadsheets and documents piling up at work. That's why Solr/Lucene and Lucid Imagination are names you need to know in tech. ... That's because of the phenomenal

<http://www.lucidimagination.com/blog/?p=2874>

[\[LUCID\] Information creation proliferation? Forbes looks to Solr and Lucid Imagination](#)

Information: we all love to make more of it, but it sure piles up. Forbes Magazine Online blogger Quentin Hardy takes on making sense of it in a nice post about Solr/Lucene and Lucid Imagination. Some soundbites: Our civilization may pride itself on the amount of information we create – more than every conversation, ever, in a couple of years; enough to jack The Library of Congress to the Moon on data-packed CD-Roms, take your superlative – but we've also made a holy mess of it. It is mostly, as they say, "unstructured," meaning as random as your last 25 emails, your tweets, and all those spreadsheets and documents piling up at work. That's why Solr/Lucene and Lucid Imagination are names you need to know in tech. ... That's because

High Recall and High Precision with Morphological Analyzer

INTRAFIND

Suchergebnisse für "buchen"

IHRE SUCHE

DATUM FILTERN **ALLE INHALTE** HEUTE 24 STUNDEN 7 TAGE 30 TAGE

GENAUES DATUM VOM BIS

8044 ERGEBNISSE

Sortieren nach: **RELEVANZ** DATUM

Filter: **ALLE INHALTE** NUR REZENSIONEN



STUDENTENANDRANG

Willkommen in der großen Maschine Universität

... rückt, der lieber dort als in Saarbrücken studiert usw. Darum müssen die Unis die Studiengänge "über**buchen**". Zum Semesterstart merken sie dann oft, dass sie sich geirrt

haben: Das "Annahmeverhalten" war besser als geschätzt. Dann haben mehr Studierende [\[weiter...\]](#)

18.10.2011, ZEIT ONLINE



SACHBUCH

Liebe und solche Sachen

..., wird nie mehr in aller Unschuld in diesem angesagten Restaurant für einen Jahrestag der Liebe den teuren Fensterplatz **buchen** oder naiv die Reise zu zweit in den

Süden für eine individuelle Entscheidung halten, also blind sein gegenüber dem Ausleben von [\[weiter...\]](#)

15.10.2011, DIE ZEIT



ZUFLUCHTSORT DEUTSCHLAND

Das gelobte Land

... Familie schon einmal einen Kurzurlaub in Deutschland **gebucht**, auf dem Oktoberfest waren sie. Die Lebensfreude der Deutschen habe seinen Kindern gefallen, sagt er. Der Rest

habe sie eher kaltgelassen, vor allem das Essen. Er lächelt. »Aber vielleicht habe [\[weiter...\]](#)

03.10.2011, DIE ZEIT



HANDY-FAHRSCHEIN

Von der Deutschen Bahn verfolgt

... Endpunkt jeder Fahrt einzugeben. Das System berechnet den Fahrpreis und **bucht** ihn vom Konto des Nutzers ab. Selbstverständlich traut die Bahn ihren Kunden nicht und will

überprüfen, ob deren Angaben stimmen. Sie sammelt dazu Bewegungsdaten. Das ist [\[weiter...\]](#)

27.09.2011, ZEIT ONLINE



SPITZENGASTRONOMIE

Frankreichs Köche setzen Jean-François Piège auf Platz 1

... ständig **ausgebucht**. Die Köche als Leser des Branchenmagazins Le Chef konnten den ihrer Ansicht nach

besten Kollegen selbst bestimmen. Es gibt keine Liste oder Vorauswahl. Neben Piège kürten sie Alexandre Jean zum "Sommelier des Jahres" [\[weiter...\]](#)

27.09.2011, ZEIT ONLINE

High Recall and High Precision with Morphological Analyzer

INTRAFIND

Suchergebnisse für "Buch"

IHRE SUCHE

DATUM FILTERN **ALLE INHALTE** HEUTE 24 STUNDEN 7 TAGE 30 TAGE

GENAUES DATUM VOM BIS

93445 ERGEBNISSE

Sortieren nach: **RELEVANZ** DATUM

Filter: **ALLE INHALTE** NUR REZENSIONEN



POLITISCHE GEFANGENE

Magischer Realismus

... Tagen hatte der alte Report des heute 84-jährigen Schriftstellers Platz eins auf Irans Bestsellerliste erklommen, wie die News-Webseite Aftab wissen ließ. Kurz darauf war das **Buch** ausverkauft. Was kümmert die Teheraner urplötzlich das [\[weiter...\]](#)

18.10.2011, DIE ZEIT



WINKLER'S "GESCHICHTE DES WESTENS"

Das deutsche Kapitel

...Heinrich August Winkler Geschichte des Westens Die Zeit der Weltkriege 1914-1945 Politisches **Buch** C.H. Beck München 2011 1350 38 Der Westen [\[weiter...\]](#)

18.10.2011, DIE ZEIT

Wirtschaftspolitik

... Praxis bedeutet. "Es gibt keine Inflationsgefahren" (DIE ZEIT, Nr. 25/2010) Jürgen Stark, Chefvolkswirt der Europäischen Zentralbank, soircht in einem Interview über die Rettung kriselnder Staaten, die Grenzen des Lehr**buch**wissens und Merkels Sparpaket [\[weiter...\]](#)

18.10.2011, ZEIT ONLINE



THINKTANKS

Politische Vordenker gesucht

... besänftigen konnte. Cross-over: Institut Solidarische Moderne Der Name ist trügerisch. Unter einem Institut stellt man sich etwas anderes vor: ein paar schicke Räume,

Bücher-Regale, Konferenztische und viele [\[weiter...\]](#)

18.10.2011, ZEIT ONLINE



ROMAN "GRUBER GEHT"

Ein Mann, ein Krebs

... es dann eben doch: Er, das Testosteronbündel, lässt Trost und Nähe zu. Es sind vor allem die beiläufigen, absurden Momente, die das **Buch** anrührend machen. Wie Gruber, schon vom Krebs gezeichnet, in seiner bislang unberührten Edelstahlküche [\[weiter...\]](#)

18.10.2011, ZEIT ONLINE



KOLONIALGESCHICHTE

Schädel im Schrank

... Arzt und Rasseforscher Eugen Fischer hatte sie »bestellt«. Von der Exekution existieren Fotos; auch beschrieb der Kolonialarzt Wilhelm Wendland 1939 in seinem **Buch** Im Wunderland der Papuas, wie er die Köpfe nach der Erschießung abschnitt [\[weiter...\]](#)

18.10.2011, DIE ZEIT

Using Word Decomposition

INTRAFIND

Suchergebnisse für "Bundeswirtschaftsministerium"

IHRE SUCHE

DATUM FILTERN ALLE INHALTE HEUTE 24 STUNDEN 7 TAGE 30 TAGE

GENAUES DATUM von bis



LEUPHANA-UNIVERSITÄT LÜNEBURG

Der Über-Bau

11.08.2014 Die 54 Millionen Euro, die der Libeskind-Bau zunächst kosten sollte, hatte er schnell zusammen. Mit dabei: das **Bundesministerium für Wirtschaft und Technologie**, das die energieoptimierte Bauweise des Audimax fördert. Die Stadt Lüneburg, die durch die ...



ELEKTROAUTO

Der Anfang vom Ende des Ladeelends

07.04.2014 ... das **Bundesministerium für Wirtschaft und Energie**. Die Säule bei der Volksbank Mehr als ein Drittel des Gesamtbudgets von 14,2 Millionen Euro kommt vom Staat, den Rest teilen sich die anderen Konsortialmitglieder. Die ...



FREIHANDELSABKOMMEN

Mehr Wachstum durch TTIP ist ein Märchen

12.11.2014 ... die USA 16-mal und für Deutschland 7-mal so groß wie in der Studie für das **Bundeswirtschaftsministerium (BMWi)**. Die Unterschiede beruhen maßgeblich darauf, dass der Arbeitsmarkt unterschiedlich modelliert wird. In der BMWi-Studie wird ein ...



ENERGIEWENDE

Überspart

11.11.2014 ... einen Park bauen will, muss sich dann darum bewerben. Der günstigste Anbieter erhält den Zuschlag. Vor allem für die weniger finanzstarken Bürgerprojekte eine Herausforderung. Zwar beteuert das **Bundeswirtschaftsministerium**, dass auch sie zum Zuge kommen ...

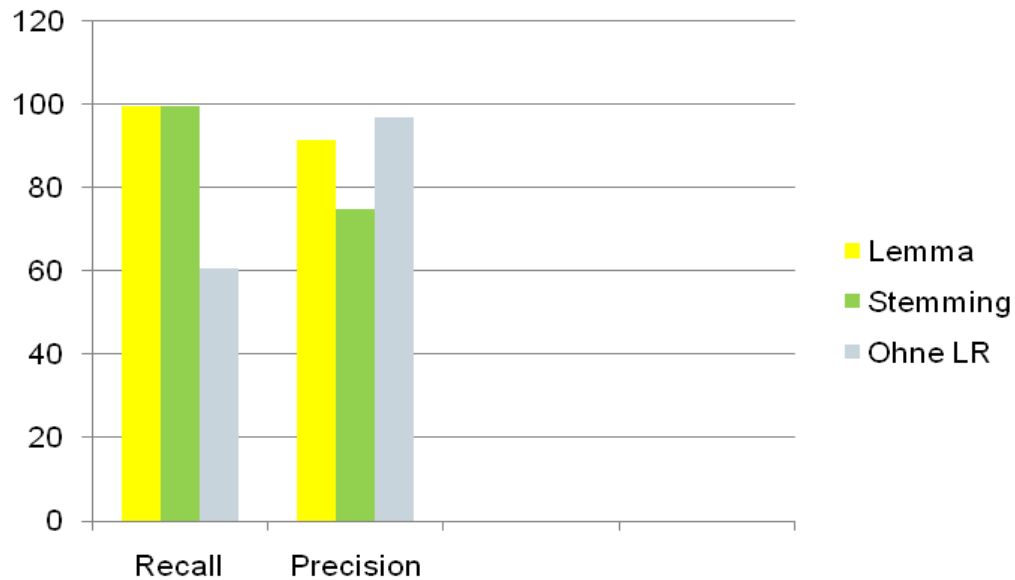


KONJUNKTUR

Deutscher Export erholt sich

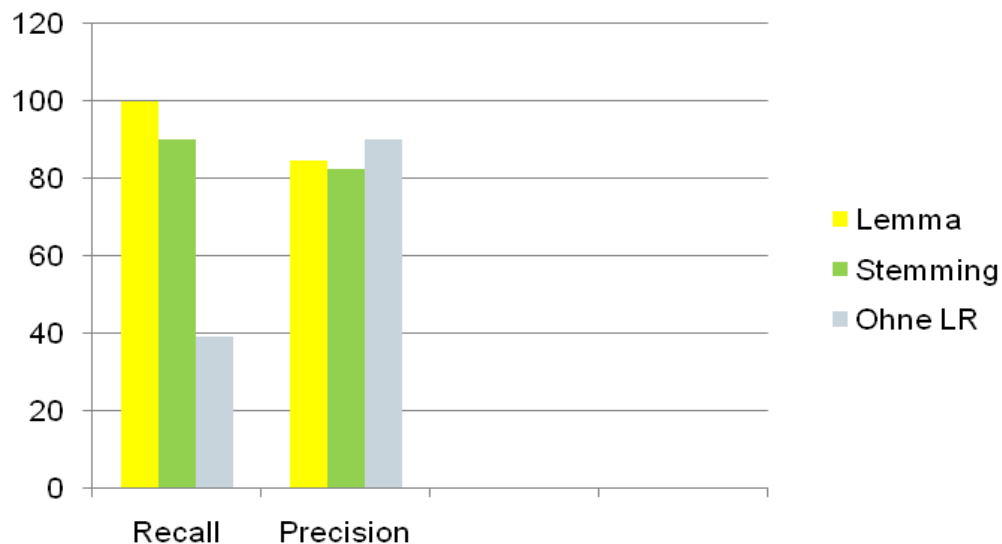
07.11.2014 ..., der maßgeblich durch die späte Lage der Sommerferien verursacht war", teilte das **Bundeswirtschaftsministerium** mit. ...

Measuring Recall and Precision



Nouns

Recall / Precision Macro Average
for 40 German nouns
no compound analysis



Verbs

Recall / Precision Macro Average
for 30 German verbs
no compound analysis

Why do we need other normalizations, besides lemmas? INTRAFIND

- ▶ MultiTermQueries: WildcardQuery, FuzzyQuery
 - ▶ No stemming, no lemmatization
 - ▶ Should work on original terms generated from Tokenizer
 - ▶ Only very simple normalizations such as
 - Diacritics: Citroën → Citroen
 - but for mou* you don't want to find mice
- ▶ Case-Sensitive
 - ▶ Sometimes case matters: MAN vs. man
- ▶ Phonetic Search (Double Metaphone):
 - ▶ Mazlum → MSLM; Muslim → MSLM
 - ▶ book → PK; books → PKS
 - ▶ Kaother Tabai → KOR TP , Kouter Tapei → KOR TP

Why do we need them simultaneously (in one field)?

INTRAFIND

- ▶ Query: “***book of Kouter Tapei***” PhraseQuery / NearQuery

Matching Document: ***books of Kaouther Tabai***

- ▶ For *book* to match *books* we need a stemmer or a lemmatizer
- ▶ For the names to match we need phonetics

- ▶ Phonetic search for query: *Mazlum*

- ▶ It leads to matches for the very frequent word Muslim
- ▶ Users want: **Give me phonetic matches for *Mazlim* but not *Muslim***
- ▶ `Mazlum=P AND NOT Muslim=E` doesn't do the job
 - No match for “*Mazlum is a member of the Muslim community in Munich*”
- ▶ `spanNot (spanOr ([body:V_mazzlim, body:F_MSLM]), body:V_muslim)`
- ▶ **New Syntax:** `<Mazlim=P BUTNOT Muslim=E>`

- ▶ We need different types of terms in one field
- ▶ Types are term properties: payloads are not a good option
- ▶ Use prefixes to distinguish them:
 - ▶ V_ for fullforms (case sensitive)
 - ▶ N_ for diacritics normalizations
 - ▶ F_ for phonetic normal forms
 - ▶ B_ for baseforms: B_Noun_book, B_Verb_fly, ...

Multilingual Search / Mixed Language Documents / Language Chunking

- ▶ Standard Approach for Multilingual Search:
 - ▶ language-specific copy of every content-field
 - ▶ language-specific analyzers for the language-specific fields
 - ▶ **Disadvantage:**
 - high configuration costs
 - no solution for mixed-language content

- ▶ **Typed Index Approach:**
 - ▶ **Language Chunker** delegates language chunks to language-specific sub-analyzer
 - ▶ Third-Party Analyzers can easily be integrated
 - ▶ Token Language is represented by type (prefix)
 - B_ for baseforms: B_EN_Noun_book, B_EN_Verb_fly, B_DE_NOUN_buch
 - ▶ **Advantage:**
 - Almost no configuration effort
 - correct treatment of mixed-language documents

In international companies people often switch between their own language and English. Content may look like that. 中国政府面对着海内外舆论长期的关注、批评与期待, 经过一年多的咨询、讨论、起草。 قصير عن تلبية الحاجة الامنية والحماية للموارد والاستثمار, المنوطة ضمن وهيبة الدولة!!.

la maladie reste silencieuse et l'on dit que l'individu est séropositif. Il présente des anticorps luttant contre le virus mais aucun signe clinique n'est apparent. 05Янв2000 АЗЕРБАЙДЖАН: НА 1 ЯНВАРЯ УСТАВНЫЙ КАПИТАЛ "АЗЕРИГАЗБАНКА" СОСТАВИЛ 8,5 МЛРД. МАНАТОВ. корр. Е.Старостина. На 1 января 2000 года уставный капитал "Азеригазбанка" Durante las grandes contiendas bélicas, los hombres de ciencias han enfrentado serios problemas médicos que originan gran número de bajas sanitarias. Stockholm - Die Suche war aufwändig, eine Woche lang durchkreuzten die schwedischen Streitkräfte die Stockholmer Schären, vor drei Wochen wurde die Aktion dann abgebrochen.

- ▶ Statistical Language Identification is considered as a solved problem:
 - ▶ (1) Cavnar, W., & Trenkle, J. (1994), (2) Dunning, T. (1994): both based on letter N-grams
 - ▶ 99% correctness for text of length ≥ 300 characters
 - ▶ Only 93% correctness for short text < 100 characters
 - ▶ Similar results reported for Tika, Google CLD on Twitter Data
- ▶ Statistical LI got more attention during the last years due to the need for reliable LI for short text:
 - ▶ Idig (infinity grams) [Nakataqni Shuyo]
 - ▶ [Tromp +] LIGA Corpus of Tweets for EN, DE, NL, FR, ES, IT (1500 tweets per language)

N-gram approach (1) with cross-validation	93.1%
LIGA-authors, special statistical approach with cross-validation	97.4%
N-gram approach (1) with lots of data	97.9%
Idig (Nakataqni Shuyo)	99.2%
Markov Chain (2) with lots of data	98.9%
Lexicon Based LI IntraFind	99.4%

Some difficult Tweets:

job tipp in der email signatur ruhig twitternamen nennen

unser musiktipp des tages louis armstrong la vie en rose eine tolle version eines klassikers musikmontag

partagez vos bons plans twitter sur les hashtags followfriday ou ff

enjoy d nom nom nom

n sarkozy pdef la parole lina brudey institutrice

en rene verbaan know how to doegoodie doegoodies

Language Identification should consist of

1. Unicode Script Identification (unicode blocks)
2. Statistical N-gram approach based on Markov Chain trained for each script separately
3. Maybe combined with an HMM to model language changes

Typed Index and Semantic Search

Named Entity Recognition

Die Beiersdorf AG ist als Dachmarke Hersteller zahlreicher Markenprodukte, darunter Marken wie Nivea, Labello, Hansaplast, Futuro, Eucerin, Florena oder Tesa. Außerdem gehören zur Kosmetiksparte die Marken Juvena, la prairie of Switzerland, 8x4, arix und die Haarpflegeprodukte von Marlies Möller. Weiterhin stellt Beiersdorf verschiedene Körperpflegeprodukte (Basis PH, Doppel Dusch, Gammon) her.

Der Hauptsitz befindet sich in Hamburg, weitere deutsche Standorte sind Baden-Baden, Berlin, Emmerich am Rhein, Hannover, Heitersheim, Offenburg und Waldheim. Der Standort in Wien wird weiter als Zentrale für Mittel- und Osteuropa ausgebaut.

Nivea ist eine geschützte Marke der Beiersdorf AG.

Die 1911 auf den Markt gekommene Hautpflegecreme NIVEA Creme ist das bekannteste Produkt der Beiersdorf AG. Den Namen leitete Oscar Troplowitz vom lateinischen Adjektiv niveus (zu nix, nivis, Schnee) ab, er bedeutet „die Schneeweiße“. Zuvor gab es bereits seit 1906 eine ebenfalls weiße Seife.

Zusammensetzung der Hautpflegecreme

Grundlage war die Entdeckung von Eucerit, einem aus Schafswollfett gewonnenen Emulgator, dem ersten Wasser-in-Öl-Emulgator. 1911 entwickelte der Besitzer von Beiersdorf, Oscar Troplowitz, eine Hautcreme in enger Zusammenarbeit mit dem Chemiker Isaac Lifschütz und dem Dermatologen Paul Gerson Unna. Im Dezember desselben Jahres kam die erste Hautcreme der Welt mit langanhaltender Wirkung auf den Markt. Die Rezeptur ist seit den Anfangstagen nahezu unverändert geblieben: unter anderem Glycerin, Panthenol, Zitronensäure, Wasser, Emulgator (Eucerit) und Duftstoffe.

<input type="checkbox"/>	Cat_ADJECTIVE
<input type="checkbox"/>	Cat_NOUN
<input type="checkbox"/>	Cat_VERB
<input checked="" type="checkbox"/>	City
<input checked="" type="checkbox"/>	Organization
<input checked="" type="checkbox"/>	Produkt
<input checked="" type="checkbox"/>	PersonName
<input type="checkbox"/>	Location
<input type="checkbox"/>	OrgCompanyTemp
<input type="checkbox"/>	Org_Company
<input checked="" type="checkbox"/>	ProduktMarke
<input checked="" type="checkbox"/>	Rohstoff
<input checked="" type="checkbox"/>	date

Entities are e.g.

- ▶ People names
- ▶ Company names
- ▶ Brands from product lists
- ▶ Technical key figures from technical data (raw materials, product types, order IDs, process numbers, eClass categories)
- ▶ Names of streets and locations
- ▶ Add Term Types for Entites such as
 - ▶ E_Person, E_Location, E_Organization
 - ▶ E_PersonName_Brown, E_Location_Munich

Semantic Search: Comparison with standard Search Engines

Question: **Wer hat Microsoft gegründet? *Who founded Microsoft?***

INTRAFIND
NL-Search



....sein Betriebssystem Windows und seine Büro-Software Office. Das Unternehmen wurde 1975 von Bill Gates und Paul Allen gegründet. Der Name Microsoft steht für Microcomputer-Software, ursprünglich...



Microsoft – Wikipedia

Das Unternehmen wurde 1975 von Bill Gates und Paul Allen **gegründet**. Juli 2004 **hat Microsoft** bekanntgegeben, dass es nach der nun erfolgten Beilegung ...

INTRAFIND
standard

Microsoft Dynamics NAV ist eine Standardsoftware für ERP-Systeme. 2002 übernahm Microsoft den dänischen Hersteller und integrierte es in seinen Geschäftsbereich Microsoft Business Solutions. Seitdem wird Microsoft Dynamics ...

Semantic Search: Comparison with standard Search Engines

Question: **Wo liegen Werke von Audi? *Where are the Audi plants?***

INTRAFIND
NL-Search



...angehört. Die Fahrzeuge der Marke Audi werden außer in den beiden deutschen Werken Ingolstadt und Neckarsulm in Győr (Ungarn), Bratislava (Slowakei), Changchun (Volksrepublik China), Brüssel...



[Audi – Wikipedia](#)

Die Fahrzeuge der Marke Audi werden außer in den beiden deutschen Werken Ingolstadt und ... 1915 wurde die „Audi Werke AG“ gegründet. Nachdem Audi 1928 in ...

[Audi und BMW: Neue Werke in Bayern - Börse - FOCUS Online](#)

Audi-Chef Franz-Josef Paefgen stößt mit seinen Werken in Ingolstadt, Neckarsulm und im ungarischen Győr an die Kapazitätsgrenzen. Wie FOCUS- MONEY erfuhr, ...

INTRAFIND
standard

... höre“ ins Lateinische „Audi“. Im Juli 1910 verließ das erste Fahrzeug mit dem Namen Audi das Zwickauer Werk. 1915 wurde die „Audi Werke AG“ gegründet. Nachdem Audi 1928 ...

Thanks for your Attention

INTRAFIND

Questions ?

By the way: Our Analyzers are available as Plugins for Lucene / Solr / ElasticSearch

Dr. Christoph Goller

Phone: +49 89 3090446-0

Fax: +49 89 3090446-29

Email: christoph.goller@intrafind.de

Web: www.intrafind.de

IntraFindSoftware AG
Landsberger Straße 368
80687 München
Germany