

## **Intelligente unternehmensweite Suche**

Suchst Du noch – oder findest Du schon?

Rutger Lörch, Leiter Vertrieb, IntraFind AG

# Kurzdarstellung IntraFind

## Firma, Kunden

## **IntraFind ist Ihr Spezialist für die Informationsgewinnung in unstrukturierten und strukturierten Unternehmensdaten.**

- ▶ Kundenbasis in Deutschland, Österreich und Schweiz
- ▶ Kunden: > 700
- ▶ Mitarbeiter: 20
- ▶ Gründung: Jahr 2000

Wir bieten unseren Kunden Softwareprodukte, Lösungen und Beratung rund um die Themen Enterprise Search, Information Access & Text Mining und verfügen über 10 Jahre Erfahrung in der Planung und Umsetzung von Suchprojekten.

Namhafte Unternehmen wie Audi, Beiersdorf, Bosch, Deutsche Post DHL, Gesellschaft für Konsumforschung, Haufe Verlagsgruppe, Krauss-Maffei Wegmann, Landesbank Berlin, Merck, Polizei Bayern, Polizei Baden-Württemberg, Rohde & Schwarz, Tech Data sowie Voith arbeiten mit unseren Softwareprodukten und Lösungen.

- ▶ Enterprise Search (Bosch, Landesbank Berlin, BHF-Bank, Eirich, GfK)
- ▶ Intranet-Suche und Onsite-Suche (Polizei Bayern, Phoenix-Contact)
- ▶ Compliance Lösungen, eDiscovery (Bundesverband deutscher Banken)
- ▶ Semantic Linking für Medienportale (Gruner+Jahr, Haufe)
- ▶ Expertenidentifikation (Deutsche Post DHL)
- ▶ Wissenslandkarten (R&D-Search Beiersdorf)
- ▶ Informationslogistiklösungen (Merck, EBV)
- ▶ Newslettermanagement / Newswatch (Merck)
- ▶ Intelligente Shopsuche

**„Suche“ geht von Simple Search bis zu Business Intelligence**

# Motivation für leistungsfähige Suche & Text Mining

- ▶ Wir erstellen, kopieren und speichern täglich Unmengen von Daten in

- ▶ Emails
- ▶ Dokumenten
- ▶ Applikationen
- ▶ Web

verteilt auf alle möglichen Geräte

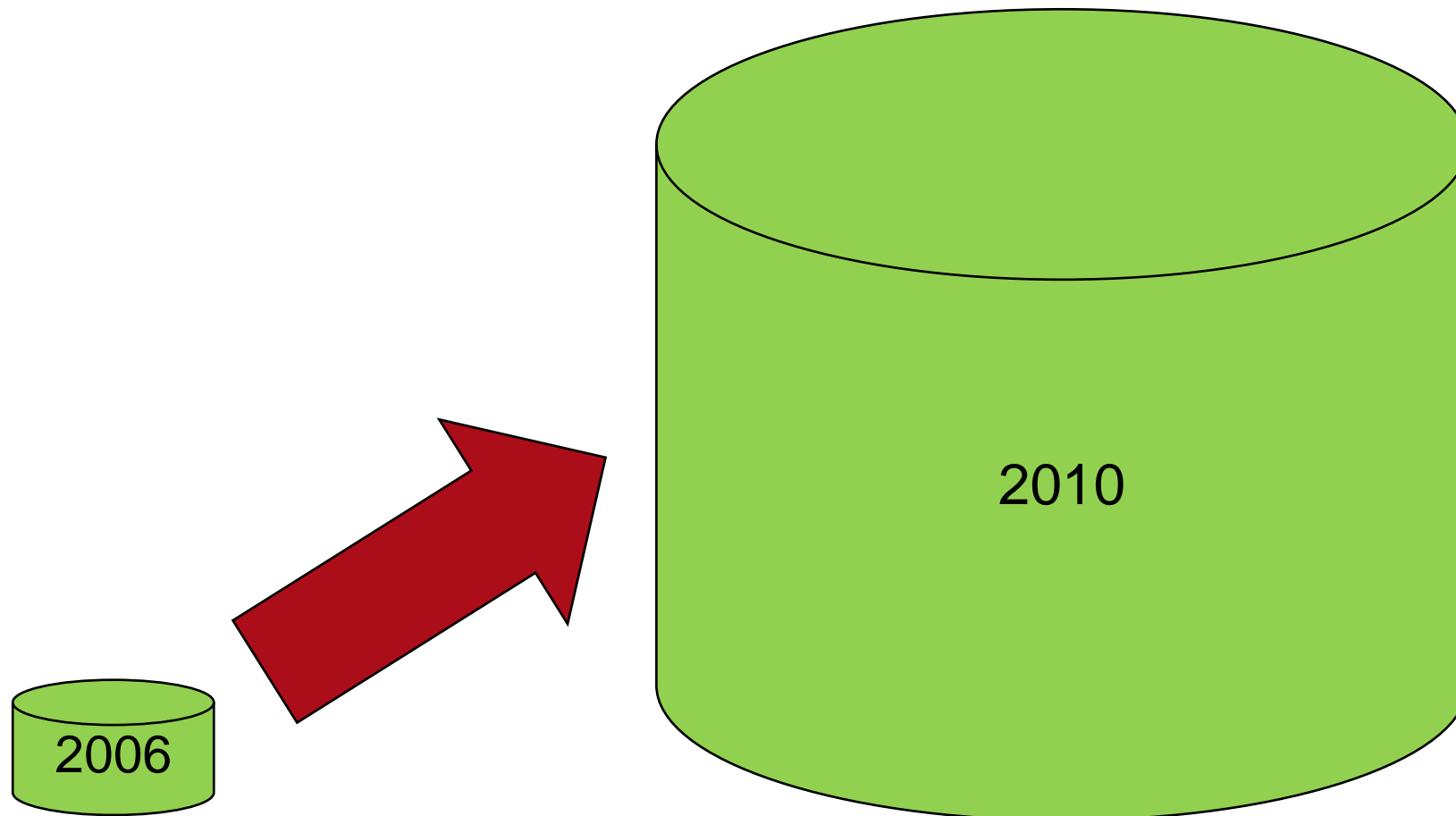
- ▶ Dann der Anruf vom Kollegen

- ▶ Wo ist denn gleich wieder...?
- ▶ Hast Du noch das Angebot von...?
- ▶ Haben wir was zum Thema...?



Das digitale Universum explodiert

- ▶ 2006 betrug die digitale Datenmenge noch **0,160 Zettabyte**
- ▶ Für 2010 wurde eine Datenmenge von **1.200 Milliarden GB** oder **1.2 Zettabyte** erwartet



- ▶ **Einsparungspotential durch deutlich kürzere Suchzeiten**
- ▶ **Effizienzsteigerung durch Integration in Geschäftsprozesse**
- ▶ **Time to Market (Produkte schneller entwickeln, relevante Patente identifizieren...)**
- ▶ **Besserer Service (selbstlernende FAQ-Systeme) / Kundenbindung**
- ▶ **Monetization (Webshops, Guided Search , Recommendation...)**
- ▶ **Ausserdem, die Mitarbeiter wollen.....**

# Kurzübersicht Enterprise Search ELO & IntraFind

2 grundsätzliche Szenarien:



ELO prof + enterprise,  
ELO 2011

intelligente Suche innerhalb ELO

# INTRAFIND

iFinder  
Enterprise Search

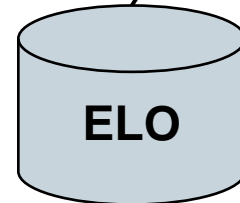
Suchfrage:

mit eigener GUI  
(Java/WebParts)

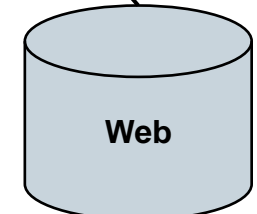
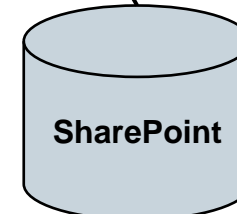
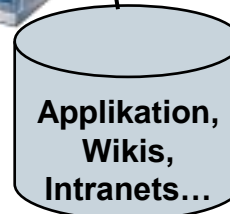
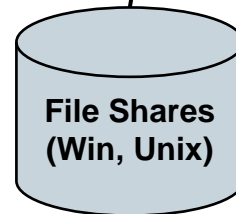


Konnektoren

SharePoint



+



weitere Quellen, die nicht in ELO enthalten sind

## Base-Paket

besteht aus folgenden Komponenten:

- ▶ Volltextsuche
- ▶ Autovervollständigung (Search-as-you-type)  
Meinten-Sie-Tippfehlerkorrekturvorschläge (Spellchecker),
- ▶ Fundstellenhighlighting in der Trefferliste
- ▶ Volle Boole'sche Suchfunktionalität,
- ▶ Dublettenerkennung,
- ▶ ConteXtoren, Facetten für geführte Suche
- ▶ Linguistik für eine Sprache,
- ▶ Synonymthesaurus DE + EN + FR

## Erweiterungspakete Entry, Advanced, Premium und Enterprise

Die Erweiterungspakete unterscheiden sich nicht funktional, sondern durch unterschiedliches Dokumentenlimit und bestehen jeweils aus

- ▶ Thesaurus-Schnittstelle (als Erweiterung des Synonymthesaurus aus dem Basispaket, für den Import eigener Thesauri oder z.B. Abkürzungslisten)
- ▶ Inhaltliche Ähnlichkeitssuche (SimFinder),
- ▶ Semantische Suche (OntologyNet) für geführte Suche & explorative Suche
- ▶ Hit-Highlighting für Hervorhebung der Suchbegriffe
- ▶ Wissenslandkarten
- ▶ Setzt Basis-Paket voraus

## Weitere linguistische Erweiterungsmodule für eine zweite oder weitere Sprachen, die das Base-Paket ergänzen

- ▶ Linguistik für weitere Sprachen

Lieferumfang: DE, EN, FR, ES, IT

- ▶ Weitere Sprachen projektspezifisch (ZH, RU, PT, CZ, SK, HU...)

# Live-Demo

- ▶ IntraFind liefert Ihnen **die beste für Deutsch bzw. Englisch verfügbare Linguistik**, die zu einer vollständigen und präzisen Trefferliste führt
- ▶ Referenzen: Haufe, Langenscheidt, Hexaglot, Gruner+Jahr, Herder Verlag, Zeit Online...

## Beispiel: Suchanfrage nach „buch“

- ▶ Trefferliste umfasst auch Treffer mit „Buches“, „Bücher“, „Büchern“, „Fachbuch“, „Sachbuches“, „Prozessmanagementhandbuch“, „Kinderbuchladen“ ...

- ▶ Erklärung:  
es werden Singular- und Pluralformen des Suchbegriffes berücksichtigt sowie gebeugte Formen oder zusammen

- ▶ Vorteil:  
es werden keine relevanten Treffer mehr übersehen

Gästebettenkapazitätsauslastung    Deutsch    **Go !**

Groß-/Kleinschreibung beachten (nur für Deutsch relevant, z.B. für **haut** und **Haut**)

Grundform	Wortart
Gästebettenkapazitätsauslastung	NOUN
<b>Komponenten</b>	
Gäste betten kapazitäts auslastung	
Gast	NOUN
Bett	NOUN
Kapazität	NOUN
Auslastung	NOUN

Klare Vorteile für den Benutzer:

Das Suchsystem ergänzt selbständig

- ▶ Pluralformen (Buch -> Bücher, Buches,...)
- ▶ gebeugte Wortformen (bringen -> gebracht, brachtest...)
- ▶ und findet auch Wortbestandteile der Suchfrage  
(Bsp.: Suche nach „Buch“ liefert auch Treffer mit „Kinderbuch“)

**Fazit:** Intelligente Unterstützung für den Benutzer -> **Vollständigkeit und Präzision**

Deutlich verbesserte Textanalyse und Textmining-Qualität für

- Assoziative Berechnungen
- Kontextanalysen
- Verlinkungsvorschläge
- Clustering

## Kontexte herstellen

Mit unserer assoziativen Suche haben wir eine Methode entwickelt, um dem Benutzer eine intelligente Auswahlmöglichkeit zu bieten, mit der er seine Trefferliste schnell und intuitiv eingrenzen kann. Die Maschine „denkt“ mit – abhängig von der Suchanfrage und dem Content, in dem gesucht wird. Die assoziative Suche berechnet die Wortwolke, es muß nichts manuell eingegeben werden!

- ▶ Zur Suchfrage „BASF“ werden die inhaltlich stark korrelierten Begriffe (auch Multiwords) aus dem Content-Bestand ermittelt.
- ▶ Die Berechnung erfolgt in Echtzeit
- ▶ Schneller Drill down in der Trefferliste -> durch Klick auf „+“ wird der ausgewählte Begriff in die Suche mit aufgenommen und eine neue Suche gestartet
- ▶ Neue Funktionalität: **Stöbern im Content**
- ▶ Anwendungsszenarien:
  - hier im Beispiel: Dynamische Navigation für User
  - Feature-Vorselektion für Themenerkennung
  - explorative Suche -> Suche nur mit den Begriffen der Wortwolke, ggf. ohne den Ursprungssuchbegriff

<input type="checkbox"/>	<b>Basf</b>	0.55	■
<input checked="" type="checkbox"/>	Ludwigshafen	0.49	■
<input checked="" type="checkbox"/>	Chemiekonzern	0.36	■
<input checked="" type="checkbox"/>	Ludwigshafener	0.25	■
<input checked="" type="checkbox"/>	Strube	0.20	■
<input checked="" type="checkbox"/>	Sondereinfluss	0.10	■
<input checked="" type="checkbox"/>	Quartal	0.08	■
<input checked="" type="checkbox"/>	Chemieriese	0.08	■
<input checked="" type="checkbox"/>	Betriebsergebnis	0.07	■
<input checked="" type="checkbox"/>	weltgrößte	0.06	■

### Ähnlichkeitssuche

- ▶ Die Ähnlichkeitssuche hilft dem Benutzer, per Knopfdruck zu einem für ihn relevanten Dokument inhaltlich sehr ähnliche Dokumente zu finden.
- ▶ Somit können z.B. in Echtzeit zu dem Artikel, den ein User gerade aufgerufen hat, inhaltlich sehr ähnliche Artikel aus dem Verlags-Content eingespielt werden.

### Themenbasierte Suche

- ▶ Die themenbasierte Suche ermöglicht Ihnen die Filterung der Trefferliste nach von Ihnen definierten Themengebieten. Die Themengebiete müssen mit Beispieldokumenten trainiert werden. Dann erkennt das Modul neue Artikel und ordnet sie den von Ihnen definierten Themengebieten zu.
- ▶ Es wird also ein Metadatum „Thema“ erzeugt, nach dem die Suche gefiltert werden kann: Zeige mir alle Dokumente, die den Begriff „Target“ enthalten und dem Thema „Monoklonale Antikörper“ angehören.
- ▶ Für die Nutzung dieser Funktion ist der Einsatz eines weiteren Produktes der IntraFind erforderlich: TopicFinder – unsere high-quality Textklassifikationslösung.
- ▶ Das automatisierte Gegenstück zur themenbasierten Suche, die trainiert werden muss, ist das Clustering der Trefferliste:

### Clustering

- ▶ Die thematische Sortierung der Trefferliste hilft dem Benutzer, sich noch schneller einen Informationsüberblick zu verschaffen. Die Trefferliste ist also nicht mehr „flach“, so wie sie es von herkömmlichen Suchmaschinen gewohnt sind, sondern thematisch gruppiert. So sorgt das Clustering für einen echten thematischen Überblick der Treffermenge.
- ▶ Die Sortierung wird wieder vollautomatisiert in Echtzeit gemacht, der Benutzer muss nichts einpflegen oder trainieren.

### Thesaurus-Anbindung

- ▶ Wir liefern unseren Kunden einen allgemeinsprachlichen Synonymthesaurus für Deutsch und Englisch mit aus, der um fachspezifische Thesauren / Wortlisten ergänzt werden kann. Der Benutzer erhält so z.B. über die Funktion „Autovervollständigung der Suchanfrage“ auch synonyme, ergänzende Begriffe zu seinem gerade eingetippten Suchbegriff vorgeschlagen, mit denen er seine Recherche erweitern kann.

## Automatische Extraktion von Entitäten / Eigennamen

- ▶ Die Erkennung von Eigennamen von Personen, Unternehmen/Organisationen und Orten bietet zahlreiche Möglichkeiten, den Benutzer zu unterstützen: So lassen sich z.B. alle Personen und Unternehmen in unstrukturierten Texten sicher und zuverlässig erkennen, das erspart den Redakteuren den Aufwand, dies über manuelle Verschlagwortung zu erreichen.
- ▶ In Kombination mit unserer semantischen Suche können Sie unbekannte Zusammenhänge aufdecken: Welche Firmen werden oft in Kontext von anderen Firmen genannt?
- ▶ Neue Zusammenhänge werden automatisiert sichtbar – so heben Sie verborgene Wissensschätze, machen dieses für Ihre Leser transparent, erhöhen die Kundenbindung und überraschen diese mit hochwertigen Informationen

## Synonyme

- ▶ Anreicherung der Suchfrage um automatisch vorgeschlagene Synonyme
- ▶ Synonymliste für Erweiterung um unternehmensspezifische Begriffe manuell oder über Import pflegbar

### 1. Beispiel:

Nach Pflege der firmenspezifischen Synonymliste werden bei Eingabe des Suchbegriffs „GSA“ auch Treffer gefunden, die den Begriff „Großbriefsortieranlage“ enthalten

### 2. Beispiel:

Standard-Synonymerweiterung: Nach Eingabe von „Etat“, werden die Synonyme „Budget“, „Haushalt“ und „Haushaltsplan“ vorgeschlagen

Etat

- Budget
- Haushalt
- Haushaltsplan

Kabinett

- Kollegium der Minister
- Nebenraum
- Regierungskabinett

Suche erweitern

Taxonomien, Akronymlisten, Thesauri, oder Wissensnetze – also generell intellektuell erfasstes Wissen lässt sich über diese Schnittstelle einbinden und im Rahmen der Suche nutzen

Wir liefern im Standard einen allgemeinsprachlichen Thesaurus für DE und EN mit aus.

# Lösungsbeschreibung: Ablagestruktur ConteXtor

### IF Ablagestruktur Contextor

- ▶ Sie können die Trefferliste schnell und einfach anhand ihrer bekannten Ablagestruktur einschränken
- ▶ Der IF Ablagestruktur ConteXtor ist sehr intuitiv zu bedienen und weist auf die absolute Anzahl der Treffer innerhalb einzelner Ordner hin

### Quicklink zum Ablageort

- ▶ Unter jedem Treffer wird ein zusätzliches Ordnersymbol eingeblendet, mit dem Sie in den jeweiligen Ordner springen können, in dem das Trefferdokument abgespeichert ist.



Content Source: Corpora | [Highlighting](#)

 <file:///gpw30001/public-1/corpora/testdaten/finder/gui/firmenprospekt.pdf> - 277KB - 29.12.2005



# Lösungsbeschreibung Wissenslandkarte

INTRAFIND

The screenshot displays the IntraFind Wissenslandkarten interface. At the top, there is a navigation bar with 'IntraFind', 'Document Center', 'News', 'Reports', 'Search', 'Sites', and 'Wissenslandkarten'. Below this, a search bar is visible with the text 'iFind'. The main content area is divided into six panels, each showing search results for a specific category:

- Author:** Search: Type in here...

Franz Kögl	365
Manuel Brunner	137
Bernhard Messer	56
Andreas Leopold	38
Herta Nagy	30
Rutger Loerch	14
IntraFind Software AG	14
Christoph Goller	6
Max Mustermann	2
Microsoft Corporation	2
- Company:** Search: in...

IntraFind Software AG	857
BOSCH Group	13
Infineon Technologies	11
Microsoft Deutschland GmbH	4
InTeCo GmbH	6
InfoChem	2
infoapps GmbH	1
Ingram Micro	1
InfoChem GmbH	1
Ihre Firma	1
- Filetype:** Search: Type in here...

pdf	34436
html	11681
class	6845
java	6719
png	4250
jar	2624
xml	2513
txt	2386
cs	2283
gif	1836
- Application:** Search: Type in here...

Microsoft Office Word	1104
Microsoft Excel	595
Microsoft Word 9.0	335
Microsoft PowerPoint	112
Microsoft Word 10.0	82
Microsoft Office PowerPoint	73
Microsoft Word 8.0	10
Microsoft Macintosh Excel	4
Microsoft PowerPoint 4.0	4
Microsoft Visio	4
- Subject:** Search: Type in here...

Search+ 2.0 und Beiersdorf	17
Enterprise Search	16
Geheimhaltungsvereinbarung	13
Findemaschine, ADAM	10
Search+ V2.0	9
iFinder Classic	8
Search+	8
Search Super+	7
iFinder	5
Intrafind	5
- Modified:** Search: Type in here...

2008	21814
2009	8008
2007	6871
2006	4515
2005	3057
2004	2511
2002	2498
2003	1508
2001	1418
2000	1387

- ▶ Start der Recherche mit 360° Blick auf IHR verfügbares Unternehmenswissen
- ▶ Absolute Anzahl der Treffer
- ▶ Frei wählbare Kategorien
- ▶ Berechtigungsprüfung !
- ▶ Performante Darstellung durch Nutzung von Ajax Technologien
- ▶ Verknüpfung zwischen Volltextsuche und Wissenslandkarten
- ▶ Intuitive Bedienung durch Guided Navigation
- ▶ Keine Schreibfehler, keine leeren Trefferlisten
- ▶ Bessere Unterstützung in Rechercheprozessen
- ▶ Finden ohne zu suchen

## Rutger Lörch Leiter Vertrieb

Tel: +49 (89) 3090446-14

Fax: +49 (89) 3090446-29

E-Mail: [rutger.loerch@intrafind.de](mailto:rutger.loerch@intrafind.de)

Web: [www.intrafind.de](http://www.intrafind.de)

IntraFind Software AG  
Fraunhofer Straße 15  
82152 Planegg