



MINDBREEZE INSPIRE

WWW.MINDBREEZE.COM

ABOUT ME



Patrick Geschwentner
Customer Success Manager

.....



Mindbreeze GmbH

Honauerstrasse 2

4020 Linz, Austria

www.mindbreeze.com

t: +43.732.606162-620

f: +43.732.606162-609

e: patrick.geschwentner@mindbreeze.com

AND YOU?

- What's your name, where are you coming from?
- What is your profession?
- You have experience with search?
- You have experience with coding, administrating, Linux,...?
- You want to share some of your interests/hobbies?

Agenda

- Web Connector
- Search
- Semantics
- Search Apps
- 360 Degree Workshop

Information and Training documentation

- All the training information will be put into a secure Fabasoft Cloud Teamroom



- Mindbreeze InSpire Training

Important Sources for you

- [Mindbreeze Website](#)
- [Mindbreeze Help Site](#)
- [Mindbreeze FAQ](#)
- [Mindbreeze Developer](#)
- Social Media
 - [Twitter](#)
 - [Facebook](#)

Web connector

Choose which pages to index

- Crawling Root(s)
- URL Exclude Pattern
 - Pages are not downloaded
 - No link following
 - Example: Javascript, CSS, ...
- Exclude URLs from Filtering
 - Pages are downloaded
 - Link following
 - Examples: product listings, blog categories, ...
- Exclude Tags From Content:
 - Remove elements from content
 - Examples:
 - Header, Footer, ...
 - if content area cannot be matched

Sitemap

- Allows for efficient delta indexing
- Short index runs
- Short interval

Metadata

- XPath + Entity Recognition
- XPath examples:
 - `//div[@id='content']`
 - `//div[@class='breadcrumbs']//a`
 - `//div[contains(@class, 'breadcrumbs')//a`
- Test in Chromium Browser using `$x("<XPath>")`
 - **Warning:** the browser executes JavaScript, the Filter Plugin does not

Metadata

- **Title:** only the first matching element is used
 - In document order
 - Default: h1, title, file name
- **Content:**
 - Every matching element

Metadata

- **mesthumbnailurl** => Type: Url, for instnace //img/@src
- **breadcrumbs**
- **mes:date** => Typ: Date
 - Format: 3. Jänner 2015 => d. MMMM yyyy&de-AT
 - Format: 3. Januar 2015 => d. MMMM yyyy&de-DE
 - Java SimpleDateFormat

Best Practices

- Exclude elements that are contained on every page
 - Menus
 - Footer
 - Header
- Use breadcrumbs
 - If they are implemented
 - Otherwise active menu entries

Training: Configuration

- **Crawling Root:** <https://www.udacity.com/sitemap.xml>

Training: Configuration

- **Name:** time
- **XPath:** `//li[h6[contains(text(),'Time')]]//h5`
- **Format:** String

Training: Configuration

- **Name:** classroomopens
- **XPath:** `//li[h6[contains(text(),'Classroom')]]//h5`
- **Format:** Date
- **Format Options:** MMMM dd, yyyy

Training: Configuration

- **Name:** prerequisites
- **XPath:** //li[h6[contains(text(),'Prerequisites')]]//h5
- **Format:** String

Training: Configuration

- **Name:** salary
- **XPath:** `//li[h6[contains(text(),'Salary')]]//h5`
- **Format:** String

Training: Configuration

- **Name:** mesthumbnailurl
- **XPath:** `//div[contains(@class,'features')]//img[contains(@src,'png') or contains(@src,'jpg')]|//div[contains(@class,'contain')]/*/img[contains(@src,'png') or contains(@src,'jpg')]/@src`
- **Format:** Url

Google on/of

<!--googleoff: all-->

I do not want to index this content

<!--googleon: all-->

Thats what I want

<!--googleoff: all-->

Please Ignore the rest

Web Connector: Extract the main content of HTML sites

Crawling sites such as news sites often indexes useless content like menus or footers.
Now you can enable an alternative extraction mode that uses heuristics to detect only useful main content.

→ You do not need to define extensive XPath rules anymore

Documentation : <https://help.mindbreeze.com/en/?topic=doc/Configuration---Web-Connector/index.htm#extract-main-content-with-alternative-filter-mode>

Web Connector: Extract the main content of HTML sites

Global Filter Plugin Properties

☐ FilterPlugin.JerichoWithThumbnails



Custom Plugin Properties

priority

20000



Add

Setup

Use Boilerpipe Extractor

Article

Allowed Values: Article

Ignore Empty Tags Pattern

Regular expression for matching html tags that should be removed if content is empty. (e.g. `^(ul|li|a)$`)

Search

Training

- <https://try.inspire.mindbreeze.com/apps/client>
- Login: david.porter;mindbreeze

Components of the client

- Restriction
- Search
- Facets
- Context actions
- Notifications

Query language

Query Language

- https://www.mindbreeze.com/Cheat_Sheet.html
- Case insensitive
- **Sof**
 - Starts of words with 3 letters
 - Whole words
 - Part of title with 3 letters
- Hits: **soft**, **Software**, im Titel auch: Micro**soft**

Query Language

- **mindbreeze ins**
 - Both terms must occur
 - Order does not matter
- Hits: **Mindbreeze**... **InS**pire, **ins**urance ... **Mindbreeze**

Query Language

- Quotes: **“mindbreeze inspire”**

Query for exact words and phrases

Hits: The European Enterprise Search Appliance | **Mindbreeze InSpire**

- **blog extension:html**

All html documents that contain blog

Hits: <https://www.mindbreeze.com/blog/search-information-access-as-easy-as-1-2-3.html>... **blog**

Query Language

- “insurance” **AND** “classify”

Hits: Companies in the **insurance** industry don’t...and can automatically **classify** it

- (“insurance” **OR** “pattern”) **AND** “classify”

Hits:

Companies in the **insurance** industry don’t...and can automatically **classify** it

Classifying by identifying patterns

Query Language

- perfect **NEAR** service

Words must be in short distance

Hits: **Perfect** community **service**

- service **NOT** blog

Documents must not contain blog

Query Language

- (a **AND** b) **OR** (c **AND** d)

Precendence without parantheses: NEAR, NOT, AND, OR

- **content:2011 TO 2013**
- **mes:size:[1KB TO 4KB]**

Ranges with units

Query Language

- `mes:date:[2015-01-01 TO 2015-12-31]`

date range

- `(url:blog^2.0 OR Mindbreeze^0.5) AND "service"`

Boost blog matches by a factor of 2 and negative boost Mindbreeze by factor 0.5

Query Language

- **title:Integration** ...**metadata,value**

all Metadata <metadata>:<value>

Name	Metadata	Description	Availability
Name	title	Query for title	All connectors
Extension	extension	Query for extension	All connectors
Directory	directory	Query for directory	Filesystem, Exchange
Subject	subject	Query for subject	Exchange

Query Language

- **title:Integration from:bauernf extension:doc**

combinable

Name	Metadata	Description	Availability
From	from	Query for senders	Exchange
To	to	Query for recipients	Exchange
Address	url	Query for URL of a document	Web
	content	Query for content	All connectors

Client Service

- Query Engines to search
- HTTP(s) Port: 443
- URL of Help-Website: tailored page or https://www.mindbreeze.com/Cheat_Sheet.html
- Language

Find out existing Metadata

- „Export“-button at the clientservice in combination with categorydescriptors
- „documents“ Servlet at the unrestricted privileged servlets
 - Find out docid:
<https://inspiretrain:8443/index/<Indexport>/find>
 - Find out metadata:
<https://inspiretrain:8443/index/<Indexport>/documents>

How to start: Know your use cases

- Which data sources?
 - Folio
 - Microsoft Exchange
 - Filesystem
 - Share Point (2007-2013)
 - ...

Preconditions

- Service user must have full-read-access to
 - File shares
 - Microsoft Exchange
 - Microsoft SharePoint
 - Fabasoft Folio
 - ...

Preconditions

- More (Authentication,..)
 - Preparations

Mindbreeze InSpire – Search Appliance

- 500K small Appliance, big Appliance, more than one big Appliance
- 1M
- 2M
- 5M
- 10M
- 20M,...

Authorization

- AuthorizedWeb vs. Web
- CN of users or groups with sitemap or url-backreferences

Semantics

Semantics

- Entity Recognition
 - Recognizes patterns
 - Partial Index rebuild necessary (reinvervation)
 - Add custom metadata
 - Use existing metadata as foundation
 - Use CategoryDescriptor and/or Information Pairing to display
 - E.g. product numbers, order numbers, dates

Semantics

- Pattern Rules
- Examples
 - `lang=/ (de|en) / .`
 - `langInTitle=/-/ lang /$/ .`
 - If Rule Matches: `langInTitle`
 - Name: `shortlang` => add to Aggregated Metadata Keys as well
 - Value: `{{lang}}`
 - Format: String
 - In Existing Metadata: title
 - `Date=/\d\d.\d\d.\d\d\d\d/ .`
 - `DateString=/am / Date .`

Metadata Definition

- **If Rule Matches** – definition is used if Pattern Rule matches
- **Name** – Name for the metadata
- **Value** – Mustache template for constructing the value, use names of rules
 - e.g. {{Date}}
- **Format** – Data type
- **Format Options** – Date format
- **Locale** – e.g. for dates
- **In Existing Metadata** – Only apply if rule matches in existing metadata

CSV Transformation

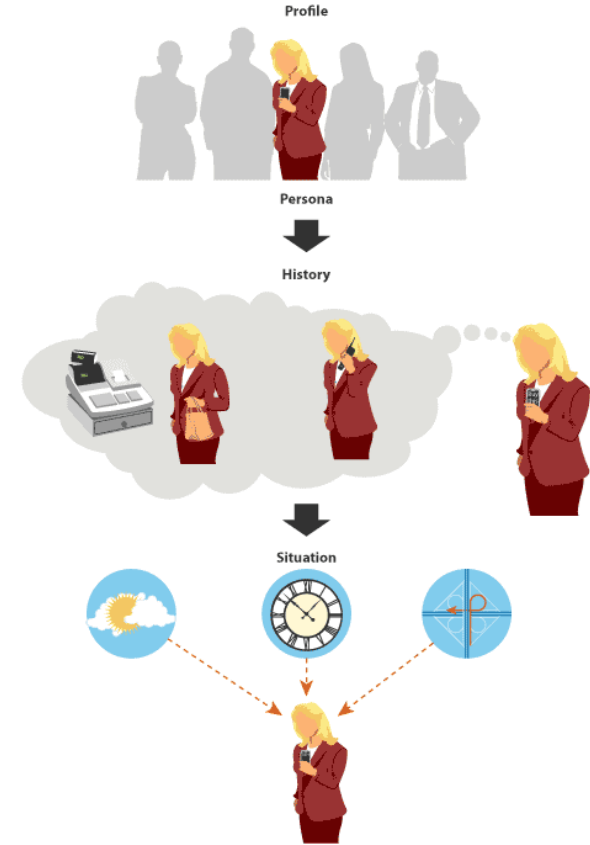
- Map values to different values
- Example
 - **CSV File Path:** /data/transform/languages.csv
 - **If Expression Matches:** {{ShortName}}
 - **In Property:** shortlang
 - **Name:** Language => add to Aggregated Metadata Keys as well
 - **Value:** {{FullName}}

Search Apps

The context is crucial

Contextualization

- Location-dependent content
- Personalized content
- Consideration of time/meetings
- Considering my Input device and the Available functions
- Consideration of form factors



Search Apps

- Tailored to users tasks
- Simple templates
- Implement search everywhere HTML is possible, e.g.
 - Microsoft SharePoint
 - Website
 - Mobile App
 - ...

Editor

- <Client Service URL>/apps/editor
- Generates Snippet
 - Copy & Paste
 - Adaptable
 - Save

Mustache Templates

Highlighting

`{{{xyz}}}` display property xyz including highlighting, try to find best representation

`{{xyz}}` display property xyz without highlighting

If

`{{#xyz?}}...{{/xyz?}}` if xyz defined

`{{^xyz?}}...{{/xyz?}}` if xyz not defined

`{{#xyz.abc?}}...{{/xyz.abc?}}` if the value of xzy is abc

Loops

```
{{#xyz}}...{{/xyz}} iterate xyz
```

Mustache

`{{xyz}}` display property xyz including highlighting, try to find best representation

`{xyz}` display property xyz without highlighting

`{{#xyz?}}...{{/xyz?}}` if xyz defined

`{{^xyz?}}...{{/xyz?}}` if xyz not defined

`{{#xyz.abc?}}...{{/xyz.abc?}}` if the value of xzy is abc

`{{#xyz}}...{{/xyz}}` iterate xyz

SearchApp Styling

CSS Framework

- Bootstrap 2.2.2 <http://getbootstrap.com/2.2.2/>

Main Structure

<http://getbootstrap.com/2.2.2/scaffolding.html#fluidGridSystem>

.container .row-fluid

.span9

.span3

Search Field

form.search-field

input

button

Results

media object component (<http://getbootstrap.com/2.2.2/components.html#media>):

.media

.media-heading

h3

.media-object

.media-body

Filters

Stacked pills-component (<http://getbootstrap.com/2.2.2/components.html#navs>):

.nav.nav-stacked.nav-pills

The header of each filter is an h3 element.

Collapsible

- ARIA role tablist + multiselectable (<http://www.w3.org/TR/wai-aria/roles#tablist>)
- Title collapsed

`.mb-role[role=tablist][aria-multiselectable=true] [role=tab]`

- Title opened

`.mb-role[role=tablist][aria-multiselectable=true] [role=tab][aria-expanded=true]`

Tabs

Basic Navbar component (<http://getbootstrap.com/2.2.2/components.html#navbar>)

.navbar .nav

Header

Basic Navbar component (<http://getbootstrap.com/2.2.2/components.html#navbar>)

.mb-header.navbar.navbar-inverse .mb-top-menu .nav

.mb-computing is added while searching

```
.mb-progress-indicator {
```

```
  display: none;
```

```
}
```

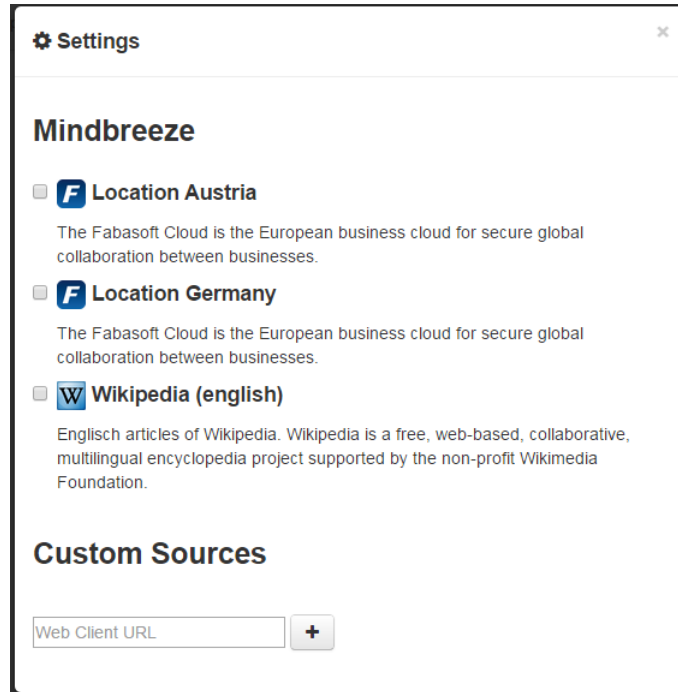
```
.mb-computing .mb-progress-indicator {
```

```
  display: block;
```

```
}
```

SearchApp Widgets

applists










breadcrumb

Refinement

☒ Living Room

collect








Collected Results

-  Manual_MLPA_DemoData 
-  24_Demonstrator_Talend.pdf 
-  Back.png 
-  Empty

constraints

Refinement






- ☒ All
- ☐  E-Mail
- ☐  Attachment
- ☐  Calendar
- ☐  Contact
- ☐  Task
- ☐  Note
- ☐  Post

datasources

Sources



- ☒ All
- ☐  msfile
- ☐  dribbble
- ☐  Web

inputdate

VON

21.07.2015

BIS

21.07.2015

Anwenden

Abbrechen

← Jul. 2015 →

Mo	Di	Mi	Do	Fr	Sa	So
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

Jul. 2015 →

Mo	Di	Mi	Do	Fr	Sa	So
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

feedbackbutton

feedback

Activate the feedbackbutton at the clientservice

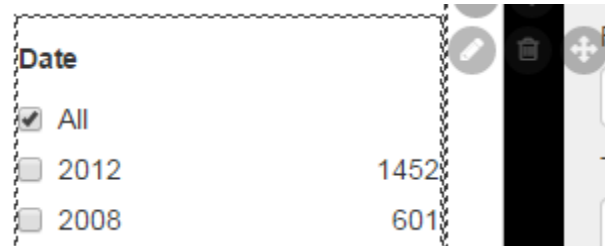
Fabasoft app.telemetry Web API URL

<http://localhost/web.telemetry>

filteredfacets

Extension		⌵
<input checked="" type="checkbox"/> All		
<input type="checkbox"/> png	2634	
<input type="checkbox"/> jpg	821	
<input type="checkbox"/> gif	564	
<input type="checkbox"/> pdf	490	
<input type="checkbox"/> docx	309	
<input type="text" value="Search"/>		
Type		⌵
Datasource		⌵
Content Type		⌵
<input checked="" type="checkbox"/> All		
<input type="checkbox"/> Document	41	
<input type="checkbox"/> UserProfile	15	
<input type="checkbox"/> Wiki Page	5	
<input type="text" value="Search"/>		

filteredfacet



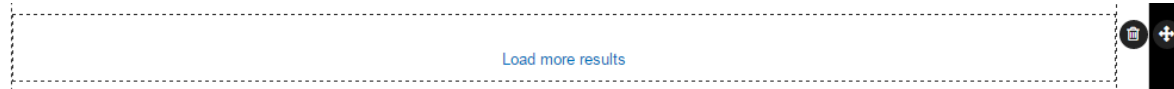
The screenshot shows a 'filteredfacet' interface. It features a table with three columns: 'Date', a count column, and an actions column. The 'Date' column has three rows: 'All' (checked), '2012', and '2008'. The count column shows '1452' for '2012' and '601' for '2008'. The actions column contains icons for edit, delete, and add. The table is enclosed in a dashed border.

Date		
<input checked="" type="checkbox"/> All		
<input type="checkbox"/> 2012	1452	
<input type="checkbox"/> 2008	601	

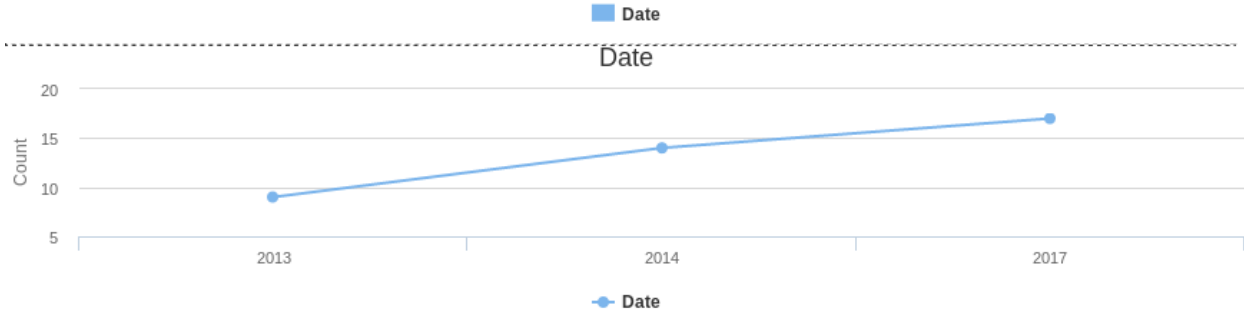
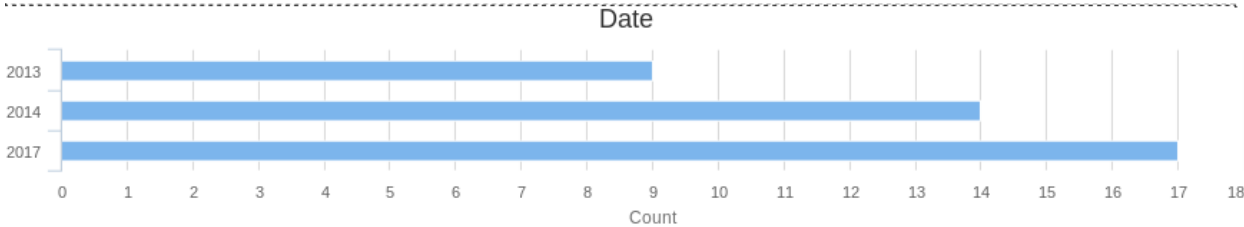
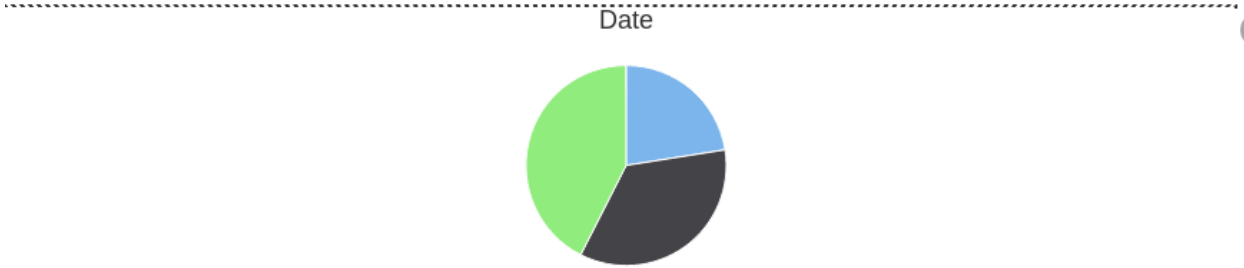
map



pagination



charts



preview



Datum: 22.10.13 09:42

Die futurezone.at zählt mit über 1 Million Visits pro Monat zu den führenden Technologie-Newsportalen in Österreich. Themen aus Politik, Wirtschaft und IT werden kompakt aufbereitet und den Leserinnen und Lesern zur Verfügung gestellt. Die Inhalte ändern sich mehrmals am Tag oft sogar mehrmals pro Stunde. Permanente Aktualität der Suchergebnisse ist damit ein wesentlicher Erfolgsfaktor. Um sich von den Mitbewerbern mit einer professionellen Suche abzuheben, wurden marktführende Suchlösungen evaluiert. Es wurde eine Lösung gesucht, die nicht nur benutzerfreundlich ist, sondern den Leserinnen und Lesern der futurezone.at auch einen Mehrwert im Vergleich zur eingesetzten Suche bietet.

Aufgabenstellung

Das Portal nutzte ursprünglich Google als Suche auf der Website. Diese lieferte zwar Treffer aber oft ohne erwarteten Mehrwert für den Leser. Die einfache Auflistung von Suchergebnissen war der futurezone.at als Leserservice zu wenig. Daher entschied das Management Mindbreeze InSite zu implementieren, um den Website-Besuchern einen Mehrwert bieten zu können und den Leserservice zu steigern.

„Als österreichisches Technologie-Portal

freut es mich, eine österreichische Suche

auf der futurezone integrieren zu können.

Es passt perfekt zu unserem Konzept,

vor allem österreichische Ideen und

Firmen vor den Vorhang zu bitten. Mit

Mindbreeze haben wir nun einen der

führenden Suchmaschinen-Entwickler als

results



Mindbreeze 3.0.pdf

[Open](#) - [Download](#) - [Preview](#) - [Collect](#)

1/29/08 9:23 PM - Directory: Handbuch - Folder - Size: 1.29 MB - Extension: pdf

www.mindbreeze.com KNOWLEDGE IN...meet this challenge Mindbreeze Enterprise Search enables...software solution. The Mindbreeze Enterprise Search Platform



MINDBREEZE.doc

[Open](#) - [Download](#) - [Preview](#) - [Collect](#)

2/26/08 4:56 PM - Directory: Marketing - Size: 46 kB - Extension: doc - Title: MINDBREEZE – KNOWLEDGE IN A MATTER... -

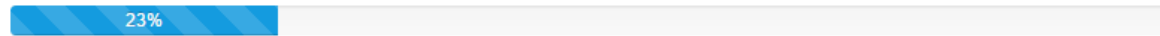
Author: Sabrina - Version: 3 - Pages: 5 - Words: 1594 - Characters: 10047

MINDBREEZE – KNOWLEDGE IN...zu meistern, unterstützt **Mindbreeze** Enterprise Search Firmen...zu organisieren. Die **Mindbreeze** Enterprise Search Plattform

Mustache example: result

```
<script type="text/x-mustache-template">  
  <h3>{{{icon}}} {{{title}}}</h3>  
  <p>  
    {{{content}}}  
  </p>  
</script>
```

resultsprogress



savedsearches

Gespeicherte Suchen

Speichern Sie Ihre Suche mit  .

[Reports letzte Woche](#)



[Meine Dokumente](#)



searchform



searchinfo

No results found

Did you mean: **pdf**

sortswitch

Sortieren nach

Relevanz



stack

20 Ergebnisse

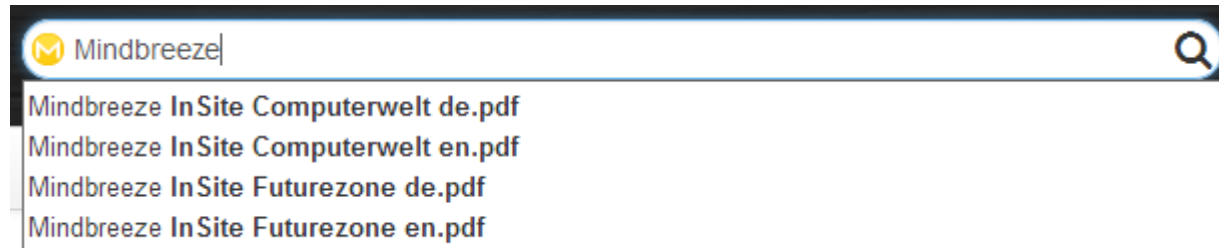
PDF-Dokumente

Vorschau **Mindbreeze InSite Computerwelt de.pdf**

 Schließen



suggest



customized suggests

Suggest Settings		
Enable Suggesting of Similar Queries	<input type="checkbox"/>	Enable suggestions of similar queries from the App.Telemetry database.
Enable Suggesting from CSV File	<input type="checkbox"/>	Enable Suggestions of Concepts from a CSV File.
Enable suggestions of Tabs	<input checked="" type="checkbox"/>	Enable Suggestions of Tabs to search for the current search term in all available tabs.
Enable Suggestions of Document Property	<input checked="" type="checkbox"/>	Enable Suggestions of Document Property to provide suggestions from metadata.
Enable Suggestions of Recent Queries	<input type="checkbox"/>	Enable Suggestions of Recent Queries to provide suggestions from the last searches.

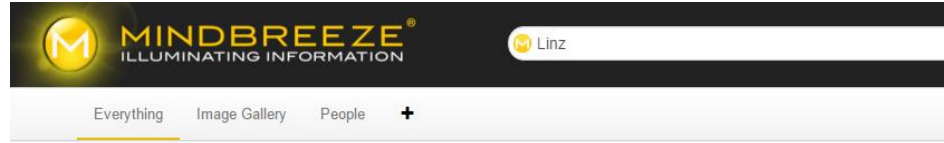
tabs

Alles Documents Wikipedia (Deutsch)

timeline

	● Mindbreeze InSpire Treibacher de.pdf	
	● Fabasoft Mindbreeze Enterprise JBA de.pdf	
	● Fabasoft Mindbreeze Enterprise JBA en.pdf	
	● Mindbreeze InSite Futurezone de.pdf	
	● Mindbreeze InSite Futurezone en.pdf	
	● Mindbreeze InSpire Ferngas OOE de.pdf	
	● Mindbreeze InSpire Ferngas OOE en.pdf	
	● Mindbreeze InSite Computerwelt en.pdf	
Okt	Nov	Dez
2013		

suggestlist



697 Results



MES Professional Search from Linz.pdf

[Open](#) - [Download](#) - [Open Folder](#) - [Preview](#) - [Collect](#)

5/5/09 3:27 PM - Directory: Press Release - Size: 1.23 MB - File type: pdf

Linz An innovative Business Search...Gartner represents the Company from **Linz**



Top250IndustryCompanies.xlsx

[Open](#) - [Download](#) - [Open Folder](#) - [Preview](#) - [Collect](#)

7/29/13 4:41 PM - Directory: Dokumente - Size: 65.56 kB - File type: xlsx - Author: Mustermann, Max

3 4020 **Linz** 050304 15 info...3 4020 **Linz** 0732/9000-0...6 4010 **Linz** 0732 771933-0...Trust Holding
Aktiengesellschaft **LINZ** AG für Energie...151 4021 **Linz** 0732 3400 office

Popular Searches

[Linz](#)

[Release Notes](#)

[Marketing Presentation](#)

[Software Documentation](#)

view

```
<div data-template="view" data-count="5">...</div>
```

```
<div data-template="view" data-count="5" data-  
constraint="extension:pdf">...</div>
```

Category

- **Category** – for every data source type
(EMC Documentum, Microsoft Exchange, ...)
- **CategoryInstance** – for every data source
(\\MYSERVER\SHARE, ...)
- **CategoryClass** – categories of documents
(Email, Contact, Appointment, ...)

Documentation and Support

- <http://help.mindbreeze.com>
- <https://tickets.mindbreeze.com>
- answers.mindbreeze.com
- developer.mindbreeze.com

Contact & News

- Contact
 - support@fabasoft.com /
 - info@mindbreeze.com
- News
 - blog.mindbreeze.com
 - twitter.com/mindbreeze
 - facebook.com/mindbreeze
 - youtube.com/Mindbreeze
 - answers.mindbreeze.com



Please provide Feedback

- URL in teamroom

Copyright/Disclaimer

Copyright © Mindbreeze GmbH, 4020 Linz, Austria, 2020.

All rights reserved. All hardware and software names used are trade names and/or brand names of the respective manufacturer.

Fabasoft accepts in this document no explicit or implicit responsibilities, in particular not as regards the completeness and correctness of the document.

This presentation contains forward looking statements, including information using the words “believe”, “assume”, “expect” or formulations with a similar meaning. Such forward-looking statements comprise known and unknown risks, uncertainties and other factors, that can result in the fact that the real results, development, financial situation or achievements deviate considerably from those assumed implicitly or explicitly in those statements. These factors among other things include: competition through other companies, effects or risks of new software and technology, the company's ongoing capital needs, financing costs, changes in the operating expenses, engaging and keeping of qualified employees, disadvantageous changes in the applicable fiscal law, riots, cause beyond control, acts of war and other in this presentation named factors. In connection with these uncertainties investors should not rely on those forward-looking statements. The company does not accept any responsibilities, to comply with those forward-looking statements in the future or to adjust them to future events or developments.

E&OE.