

Typical back-end work

This document gives an example of our work with the Mindbreeze appliance and what problems occur when configuring the indices (sources) for our application.

www.getabstract.com is one of those indices we pull data from. It offers e-books for a monthly fee. The webpage where you can see details for one e-book getabstract calls a summary. Of course, the search results from getabstract should include the title of the e-book.

Problem

The metadatum *title*, which the Mindbreeze appliance pulls, includes the following string:

summary of<lots of tabs and line feeds><actual title of the book>

So, for an e-book with the title


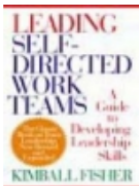
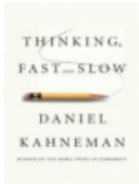
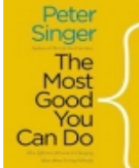
10 Ways to Have a Better Conversation

the Mindbreeze appliance pulls

Summary of

10 Ways to Have a Better Conversation

This makes the search results not look that appealing:

	<p>Summary of Is a growing middle class good for the poor?</p> <p>Open - Preview - Collect</p> <p>1/11/18 1:00 AM</p> <p>Is a growing middle class good for the poor? Summary Raj M. Desai and Homi Kharas 6 Style Recommendation Extreme poverty remains one of the most vexing problems of our time. In this intricate study, academics Raj Desai and Homi Kharas tackle the little-understood relationship between...</p>
	<p>Summary of Leading Self-Directed Work Teams</p> <p>Open - Preview - Collect</p> <p>1/11/18 1:00 AM</p> <p>Leading Self-Directed Work Teams Summary Kimball Fisher 4 Style Recommendation Kimball Fisher provides a revised, updated look at the techniques for successfully leading self-directed work teams (SDWTs). He draws on basic principles described in his earlier book of the same name. S...</p>
	<p>Summary of Thinking, Fast and Slow</p> <p>Open - Preview - Collect</p> <p>1/11/18 1:00 AM</p> <p>Thinking Fast and Slow AudioBook Daniel Kahneman This book is very good null 4 months ago Love it. Good and mind-blowing. Ola Hashem 6 months ago I bought the book an...</p>
	<p>Summary of The Most Good You Can Do</p> <p>Open - Preview - Collect</p> <p>1/11/18 1:00 AM</p> <p>The Most Good You Can Do Summary Peter Singer 7 Style Recommendation "Effective altruism" pioneer Peter Singer advocates taking a pragmatic approach to giving. To reduce suffering and save lives, do your research and place your money with organizations where it will do the most...</p>

The goal is to only have the actual title as the *title* metadatum, so we had to get rid of the „summary of“ (or for german e-books „Zusammenfassung von“).

Solution with Entity Recognition

With Entity Recognition, Mindbreeze offers a way to generate new metadata based on existing metadata using regular expressions (they call it *rules*).

First, we have to define a rule that matches the *title* that the appliance pulled, so we can work with it, we called it *TitleRule*

```
TitleRule = /((Summary of)|(Zusammenfassung von))(\t|\n)+/TitleKidRule//.
```

In this rule, we reference another rule called *TitleKidRule*, which matches that part of *title* we actually want. We don't know how exactly a book title will look, so the rule looks as follows.

```
TitleKidRule = /[-\\!A-Za-z0-9äüö]+/.
```

So technically, TitleRule looks like this

```
/((Summary of)|(Zusammenfassung von))(\t|\n)+[-\\!A-Za-z0-9äüö]+/.
```

and it matches any english or german *title* metadatum.

Just in case you are wondering: Yes, regular expressions look complicated and it took us quite a while to get these ones right.

The process

First we create a new metadatum *booktitle*, which should not include *summary of* or *Zusammenfassung von*.

If TitleRule applies – which is always the case for english and german books – the new metadatum *booktitle* is set to what TitleKidRule matched, that way we can extract only the part that we need.

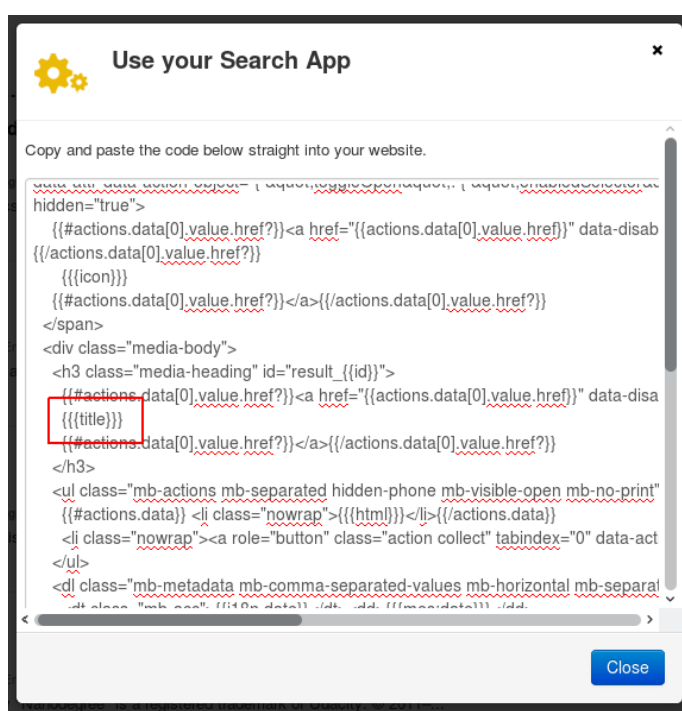
Now, every e-book has an extra metadatum, which contains its actual title:

Name	bookname
Summary of The Bully at Work	The Bully at Work
Zusammenfassung von Arbeite klüger – nicht härter!	Arbeite klüger – nicht härter!
Zusammenfassung von Der Störfall	Der Störfall
Summary of 10 Ways to Have a Better Conversation	10 Ways to Have a Better Conversation
Zusammenfassung von Green Marketing	Green Marketing
Zusammenfassung von Lohnt sich Wissensarbeit noch?	Lohnt sich Wissensarbeit noch
Zusammenfassung von Die Kunst des klaren Denkens	Die Kunst des klaren Denkens
Summary of Transforming Cities with Technology	Transforming Cities with Technology
Summary of Conversational Intelligence	Conversational Intelligence
Zusammenfassung von Krisenstaat Türkei	Krisenstaat Türkei

Upcoming Work for the front-end team

using the new metadatum

Per default, the appliance uses the *title* metadatum and displays it within the search results. To change that, when exporting the results widget from Mindbreeze UI Designer, the code has to be adjusted:



{{title}}

has to be changed to

{{booktitle}}

and it should display correctly

However, we only want to use *bookname* for getabstract. For all the other indices we still want the appliance to display the *title* metadatum. So we need to check if *bookname* is available and if not (for all the other sources) use *title*.

Moustache Templates

Moustache template offer us a way to check if a metadatum is existing or even if a metadatum includes a specific value. For this use-case, the former will serve us just fine.

Say we want to check if the metadatum *language* is available.

if *language* is defined

```
{{#language?}}...{{/language?}}
```

if *language* not defined

```
{{^language?}}...{{/language?}}
```

Given the fact that the front-end team has already included the search results in their webpage, they can change their code directly as shown above.