

# APPROACHES TO IMPLEMENT SEMANTIC SEARCH

Johannes Peter  
Product Owner / Architect for Search

**MediaMarktSaturn**  
Retail Concepts

# WHAT IS SEMANTIC SEARCH ?



# Success of search

- Interface of shops to brains of customers
- Wide range of usage
- Success depends on a proper understanding



# Simple keyword search



mymobile 7 without contract

<b>title</b>	<b>type</b>	<b>description</b>	<b>attribute</b>
MyMobile 7	Smartphone	... with a contract ...	contract
MyMobile 7	Smartphone	...	
Marriage without contract	DVD	...	
MyMobile 6	Smartphone	...	
Sitcom season 7	DVD	... 7 ...	
MyMobile 6	Smartphone	... with a contract ...	contract

# Identifying entities



mymobile 7 without contract

product / product group

without

certain attribute

Entity	Example
Products	mymobile 7
Attributes	contract
Product with / without attribute	mymobile 7 without contract
Product group with approximate price	mymobile under 300 euro

# Semantic search



mymobile 7 without contract

product / product group

without

certain attribute

<b>title</b>	<b>type</b>	<b>description</b>	<b>attribute</b>
MyMobile 7	smartphone	...	
MyMobile 6	smartphone	...	
MyMobile 7	smartphone	... with a contract ...	contract
MyMobile 6	smartphone	... with a contract ...	contract

# Core benefits



**Better precision**



**Better  
recommendation**



**Facilitated search  
management**

# Future perspectives



**Voice search**



**Sophisticated sales  
advisors**



**Chat bots**



# APPROACHES



**MediaMarktSaturn**  
Retail Concepts

# ONTOLOGIES & RULE COLLECTIONS



# Ontologies & rule collections

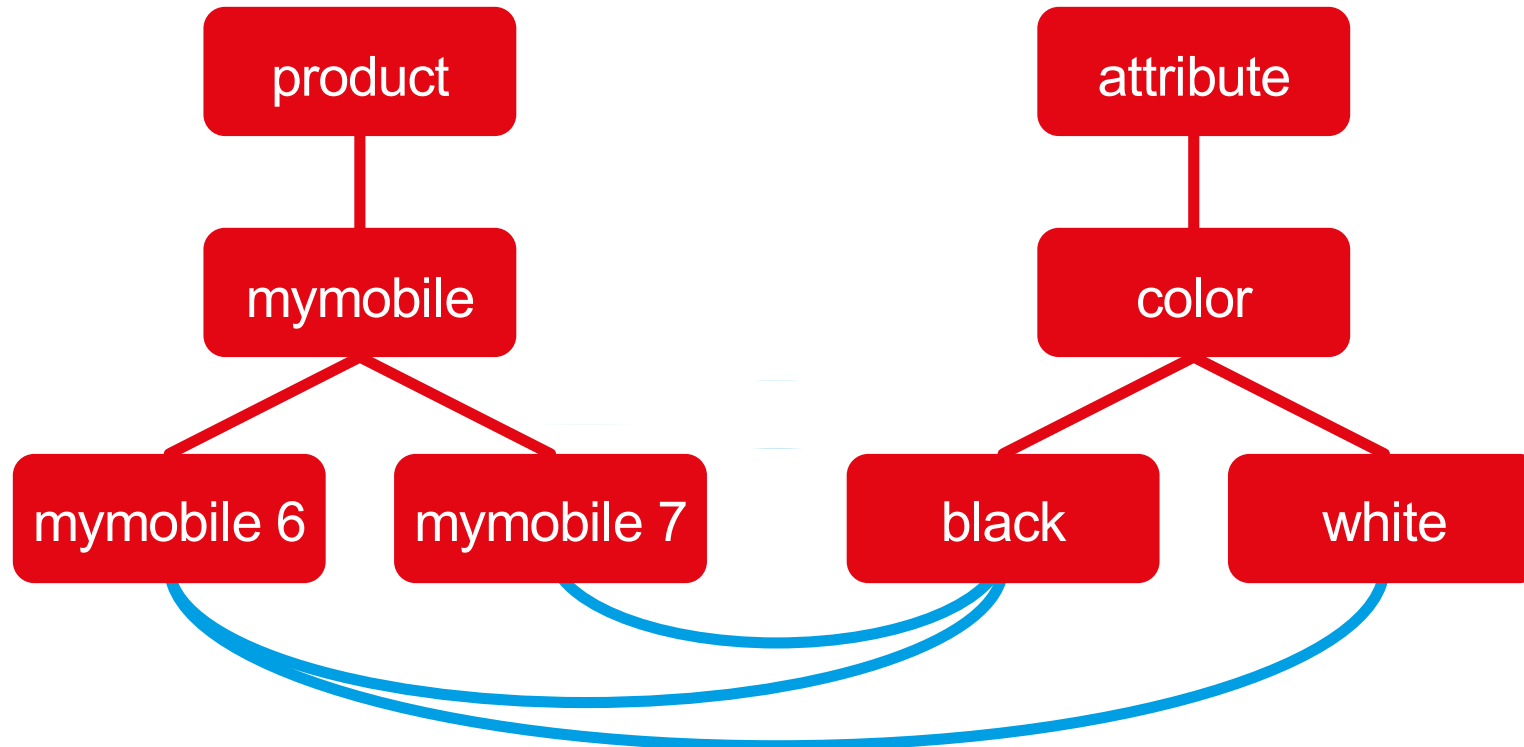


mymobile 7 without contract

Step	Example
Identify entities	<i>product(mymobile 7) without attribute(contract)</i>
Execute rules to combine entities	<i>product(mymobile 7) not(attribute(contract))</i>
Translate into search query	title:("mymobile 7") AND NOT flag:(contract)

# Ontologies

- Hierarchies of entities
- Products, attributes and relations



# Rule collections



mymobile 7 without contract

- Condition: There is the term *without* between a product and an attribute
- Action: Negate the attribute



pink dvd

- Pink: color or artist? → Disambiguation
- Condition: The term pink appears together with entities related to music or movies
- Action: Annotate the term *pink* as artist

# Implementation

- Two parts of implementation
  - Development of the application
  - Information extraction part (creation of ontologies & rule collections)
- Service for ontology extraction
  - Solr and Elasticsearch are not suitable
  - Highly scalable and performant solution with Spring Boot & Apache Lucene (using term vectors as payloads)
- Rule engine
  - Configurable rulesets
  - Routing concept

# Implementation

- Well suited for agile development
- Pieces of information can be extracted fairly independently



# Implementation

- More complex cases
  - Extract information out of product descriptions
  - Understanding of natural language



- Requires maintenance for ontologies and rule collections



# MACHINE LEARNING



**MediaMarktSaturn**  
Retail Concepts

# Machine learning



# Machine learning



<b>term</b>	mymobile	7	without	contract
<b>part of speech</b>	noun	digit	preposition	noun
<b>relation</b>	head	mymobile	contract	mymobile
<b>chunks</b>	noun phrase		noun with negation	
<b>entity</b>	product with negated attribute			

# Machine learning – NLP

- How natural is the language used for queries?
- Considering grammatical information can be complicated
- Disambiguation is very difficult for some cases

term	pink	mymobile
part of speech	adjective	noun

term	pink	dvd
part of speech	proper noun	noun

- Natural language processing:
  - "The label saw potential in Pink and offered her a contract."

# Implementation

- Established procedures from the area of natural language processing
- Libraries (e. g. spaCy) providing
  - Functionalities fairly easy to use
  - High performance
  - Customizations
- All discussed steps require their own model (training + evaluation data)
- Still highly experimental
  - Fail early?
  - Continuous delivery?

# TERM CO-OCCURRENCES



# Term co-occurrences

- Enrich documents by contextual information
- Using collaborative filters (recommendation)
- Which terms / attributes appear in the context of a product?

# Term co-occurrences



mymobile 7

title	category	color	description
MyMobile 7	MyMobile	black	Smartphone MyMobile 7 black with 128 gb
MyMobile 7	MyMobile	white	New smartphone MyMobile 7, 64 gb, white
Sitcom season 7	DVD		Season number 7 of the sitcom ...
MyMobile 6	MyMobile	black	MyMobile 6 – smartphone – 32 gb – black
MyMobile 6	MyMobile	white	MyMobile 6, smartphone black with 128 gb

- Co-occurring terms for category MyMobile:
  - Term "smartphone": 7, black, white



# Term co-occurrences



mymobile 7

<b>title</b>	<b>category</b>	<b>color</b>	<b>context</b>
MyMobile 7	MyMobile	black	6, white
MyMobile 7	MyMobile	white	6, black
MyMobile 6	MyMobile	black	7, white
MyMobile 6	MyMobile	white	7, black
Sitcom season 7	DVD		...

# Implementation

- Fairly easy to implement
- Generic
  
- Produces side effects
- Requires high data quality
  
- Only partially solves problems related to semantic search
- Not suitable for complex cases

# Conclusion

	Term co-occurrences	Ontologies + rules	Machine learning
Effort	moderate	high	high
Holistic solution	no	yes	yes
Suitable for complex cases	no	yes	yes
Maintenance effort	low	high	low
Success factors	<ul style="list-style-type: none"> <li>• High data quality</li> <li>• Agile development</li> </ul>	<ul style="list-style-type: none"> <li>• Ability of linguists</li> <li>• Quality of rules</li> <li>• Agile development</li> </ul>	<ul style="list-style-type: none"> <li>• Ability of data scientists</li> <li>• Quality of training data</li> </ul>
Risk factors	<ul style="list-style-type: none"> <li>• Side effects</li> </ul>	<ul style="list-style-type: none"> <li>• Never-ending rule-building</li> </ul>	<ul style="list-style-type: none"> <li>• Never-ending generation of training data</li> <li>• Too high expectations</li> </ul>

# THANK YOU !!



BTW: We are hiring ...

[peterj@mediamarktsaturn.com](mailto:peterj@mediamarktsaturn.com)

**MediaMarktSaturn**  
Retail Concepts