

Intelligent Work Support for Industrial Service Desks with Cognitive Services

Extracting Semantic Problem Descriptions from
Noisy and Multi-Lingual Texts

Jana Koehler, Dario Lötscher

Etienne Fux, Florian Herzog, Kai Waelti,

Hochschule Luzern - Informatik

Roland Imoberdorf, Dirk Budke

UMB AG



Customer Support is Facing Growing Complexity



«Man muss Menschen mögen»

Wer die Nummer des EKZ Kundendienstes wählt, wird ebenso unkompliziert wie kompetent betreut. Insgesamt 27 Mitarbeitende beantworten pro Jahr rund 80 000 Anrufe und 50 000 E-Mails.



Soft & Kraft

DAS KUNDENMAGAZIN DER ELEKTRIZITÄTWERKE DES KANTONS ZÜRICH AUSGABE 3/16

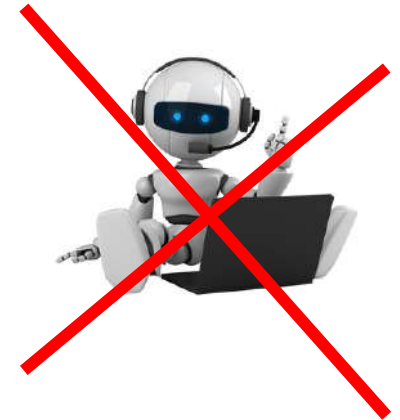
- **customized products**
- **product connectivity**

UMB Transformation
Technology

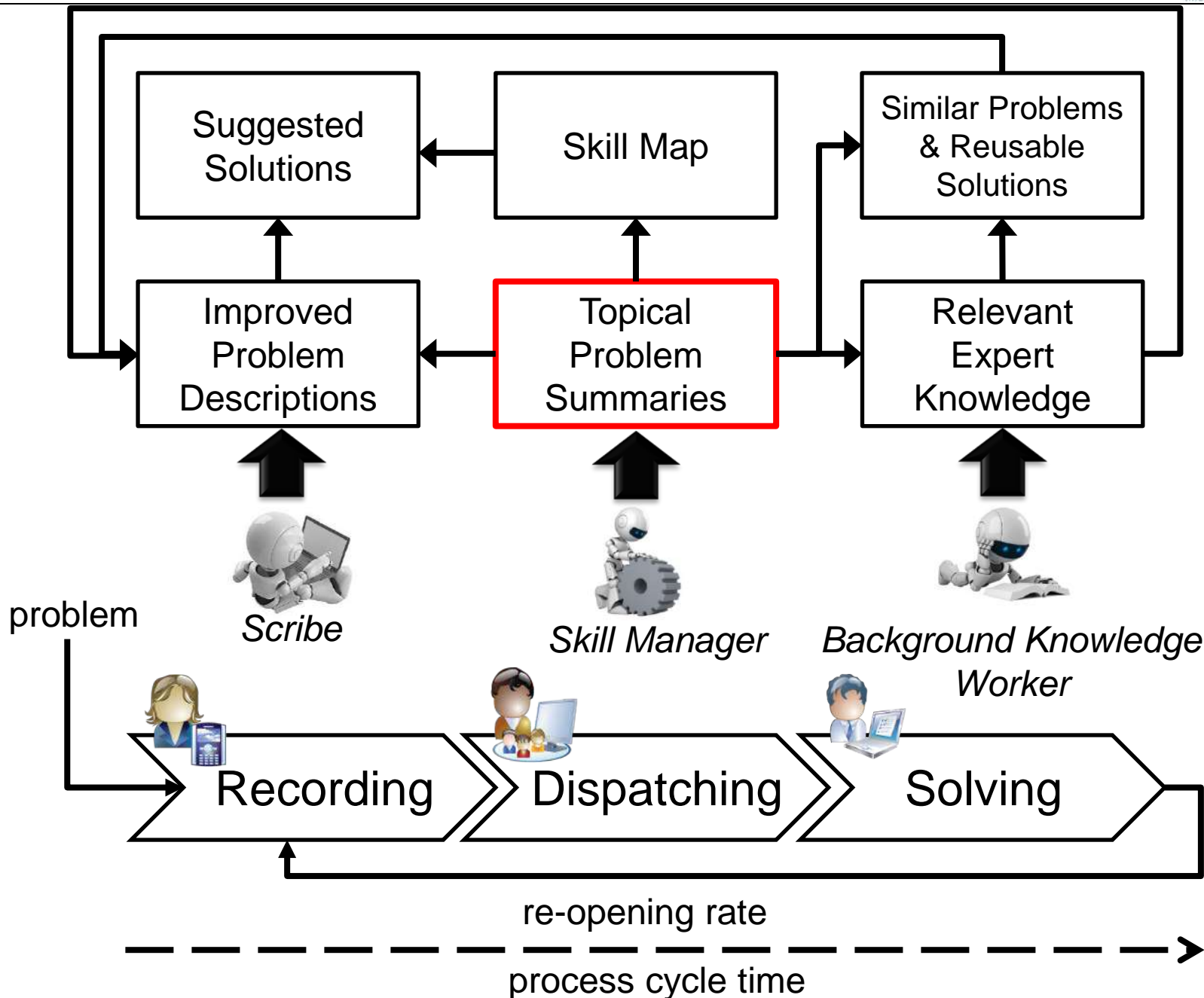
- major IT service provider
- > 100 Mio CHF revenue
- > 700 customers (UBS, IKEA,...)
- > 60.000 service requests in 2016
- double-digit growth

Employee Support vs. Automation

- Fast access to
 - best available expert
 - required information
- More personal customer contacts
- Gain knowledge from problems and solutions
 - employee training
 - improved engineering
 - optimized solution process

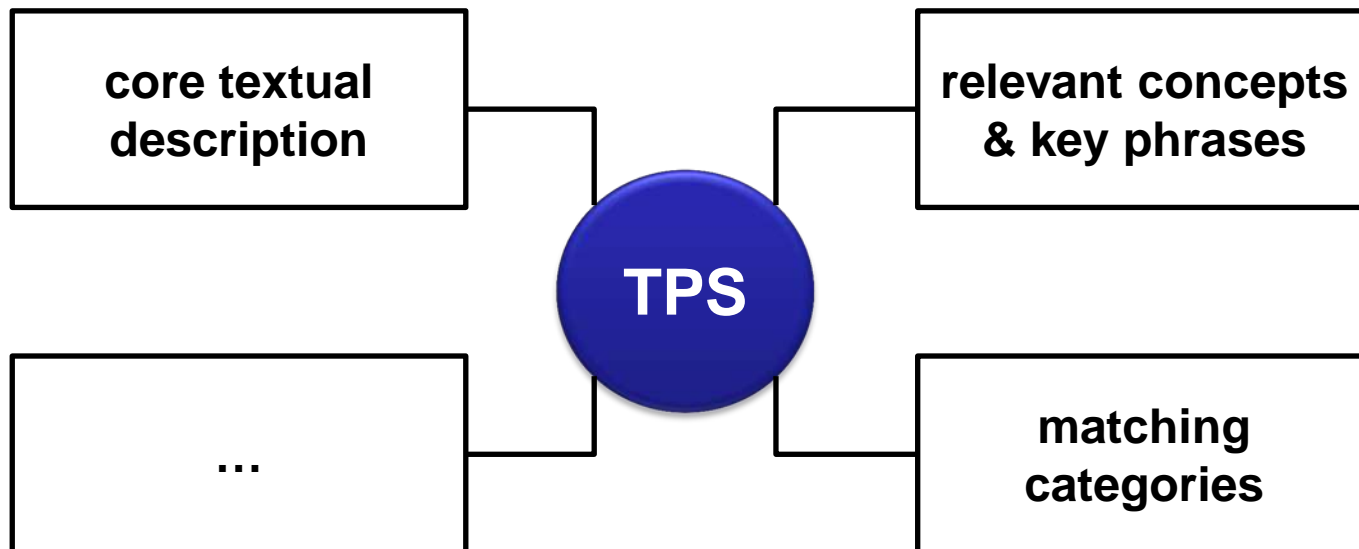


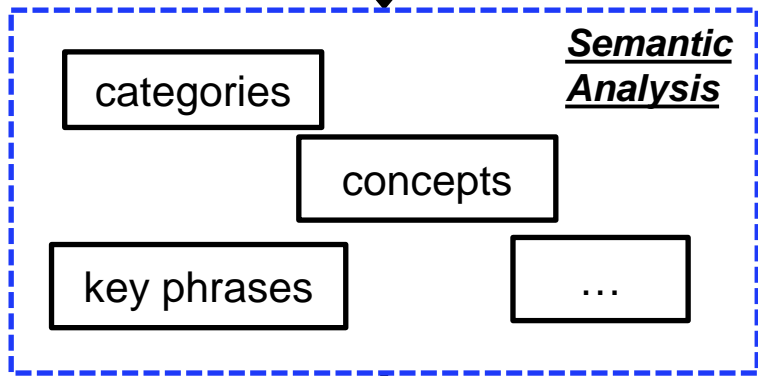
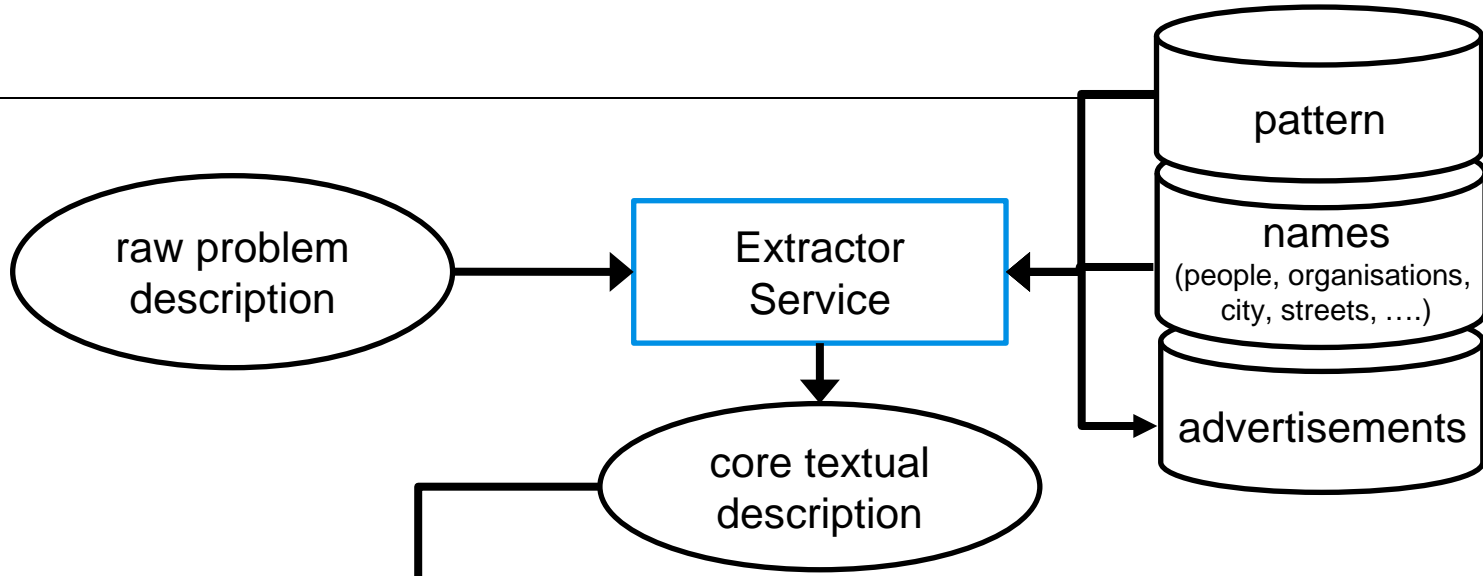
Chatbot, no thanks!



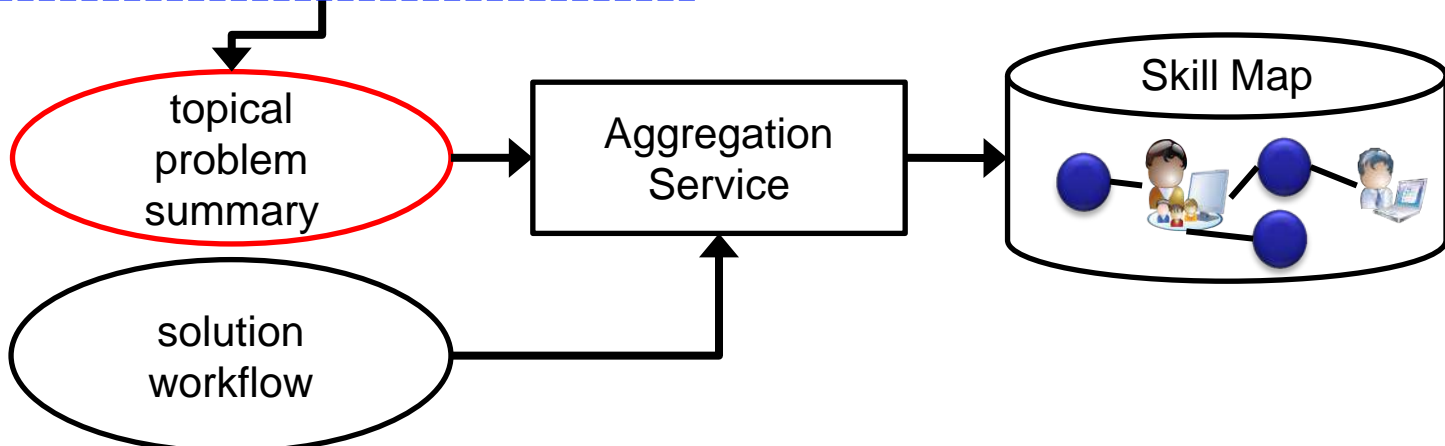
Towards Topical Problem Summaries

- A semantic description of the essence of a problem, suitable to
 - summarize and generalize problems
 - find similar problems
 - reuse solutions





The Skill Manager



Tickets

- Over 80,000 from Fall 2015 to Spring 2017
 - submitted by humans or monitoring systems
 - in 2016: approx. 25,000 submitted by humans

- Often result from email conversations

- "Technoid" language
 - technical terms
 - grammar and spelling errors
 - multiple languages
 - sensitive information - *data protection matters!*

- Solved in a team effort - solution workflow

Example - Initiating Email (mostly in German)

```
Guten Tag
Ich möchte mit X1 auf die X2 in X3 zugreifen.
Unten die Anforderungen von der IT X3.
Ich bitte um Unterstützung beim Beantworten der Fragen und Umsetzung der Anforderungen.
Bitte rufen sie mich an.
Danke und freundliche Grüsse
X4
-----
X AG
X4
mechanical engineer, dipl.Ing.HTL
O & M software engineer
Service
Xtrasse 00, Postfach 000
CH-0000 X
Phone++00(00) 000 00 00
Fax ++00 (00) 000 00 00
mail: ...@....com
web: www....com
```

- followed by over 200 lines of text in German, Finnish, English

Example - Closing Email (mostly in English)

From: X5 [mailto:....ch]
Sent: 26. huhtikuuta 2016 0:18
To: X6 [mailto:....fi]
Subject: X3 Internet Access to ROM

(1)

Dear X7,

(2)

we have no internet access anymore to the ROM PC.
Was there a change that X8 has direct access?
IP address in the ROM PC was 000.000.00.000
Our access from outside X9 was 00.000.000.0

Mit freundlichen Grüßen

(2)

X4

(5)

X AG
Xstrasse 00
0000 X
Schweiz
Tel. +00 00 000 00 00
Fax +00 00 000 00 00

(3)

This email may contain confidential and/or legally privileged information. If you have received it in error, please notify the sender immediately and delete it (together with any attachments) from your system without using or disclosing its contents for any purposes or to any other person. Many thanks for your co-operation.

(4)



Experiments with Cognitive Services



- Microsoft Text Analytics API
 - sentiment analysis, key phrase extraction
 - topic detection only for > 100 documents in a set

- IBM Watson Natural Language Understanding
 - sentiment analysis, key word, entity, concept, category detection

Sentiment Analysis

- 312 text lines including empty lines
- Microsoft TA rejects full text because of size (10240 bytes)
 - score between 0 and 1 (0.5 neutral)
- IBM NLU set to English/German
 - score between -1 and 1 (0 neutral)

	Full (312 lines)	Full Extracted (78 lines)	German Full/Extracted	English Full/Extracted
MS TA Sentiments	-	0.97	0.73 / 0.50	0.99 / 0.21
IBM Sentiments	0.01	-0.31	0.3 / 0.0	0.01 / -0.74

➤ **Better not used on our texts**

MS Key phrases / IBM Key words

	Complete Raw Text (312 lines)	Complete Extracted (78 lines)	German Raw / Extracted	English Raw / Extracted
MS TA Key phrases	-	long list of words	long list of words	long list of words
IBM NLU Key words	X3 Internet Access (0.95) mailto (0.81) H F (name) (0.70)	right IP config (0.94) adapter settings (0.92) remote connection (0.91)	41 44 277 (0.90) r@hz (0.83) s.r (name) (0.78) ----- Klopfwerksteuerung in X (0.97) Umsetzung der Anforderungen (0.80) rufen (0.54)	H F (name) (0.97) X AG (company) (0.81) ROM PC (0.77) ----- ROM (0.94) IP address (0.82) internet access (0.80)

➤ **Only partially useful for texts in English**

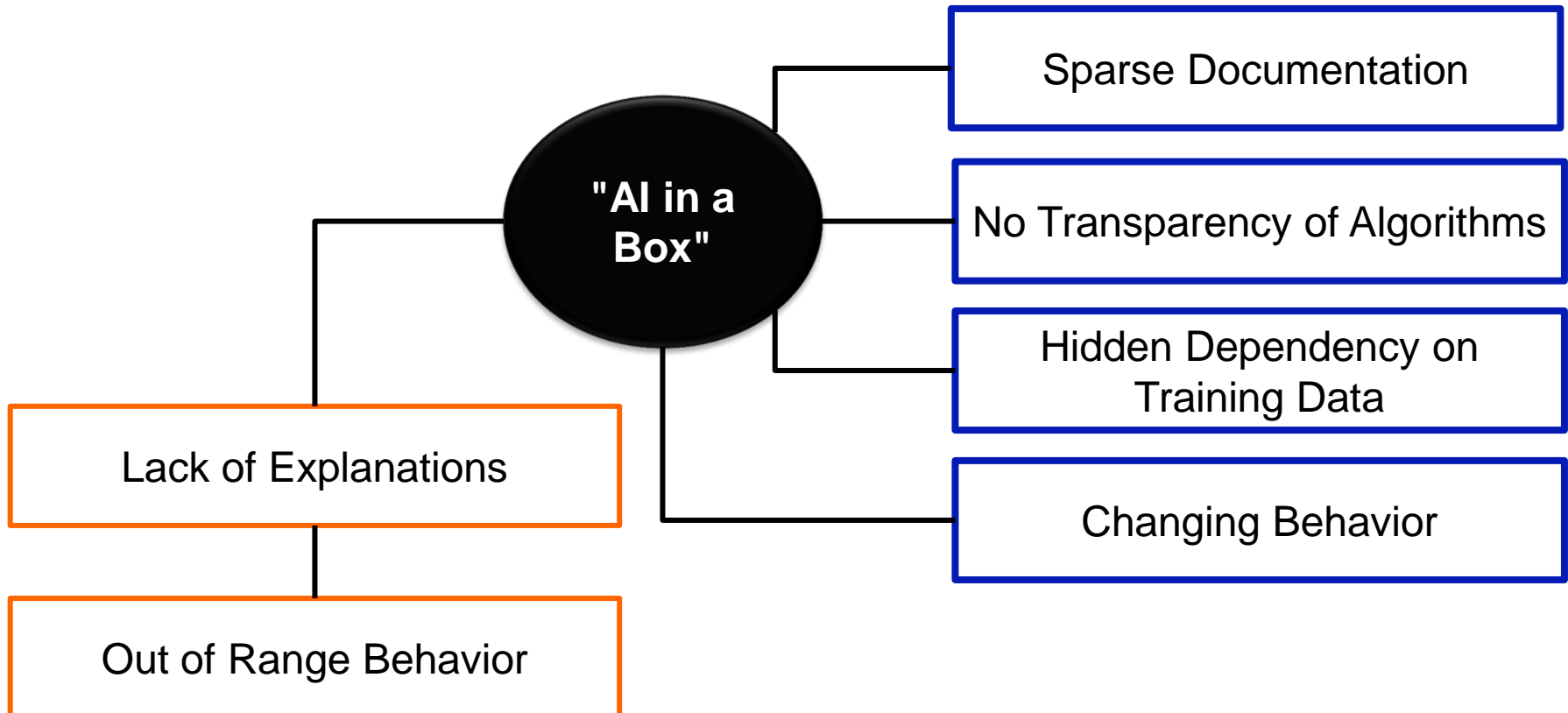


IBM NLU Categories and Concepts

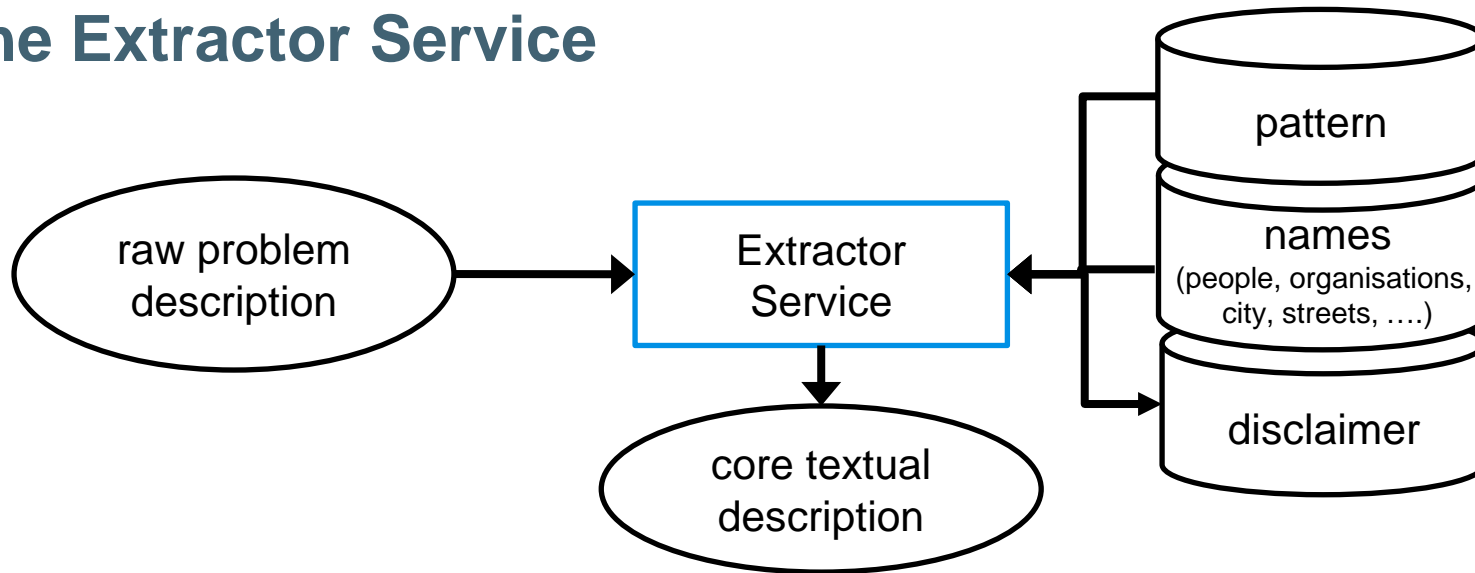
Raw (312 lines)	Raw Extracted (78 lines)	German Raw / Extracted	English Raw / Extracted
javascript (0.63) router (0.48) vpn and remote access (0.44)	router (0.58) vpn and remote access (0.57) computer (0.27)	"categories: unsupported text language",	software (0.62) / (0.39) vpn and remote access (0.53) / (0.63) computer (0.47) / 0.58)
IP address (0.95) Subnetwork (0.64) Dynamic Host Configuration Protocol (0.57)	IP address (0.97) Dynamic Host Configuration Protocol (0.78) Subnetwork (0.78)	"concepts: unsupported text language"	Internet (0.95) IP address (0.94) Internet Protocol (0.65) ----- IP address (0.98) Internet (0.93) Internet Protocol (0.70)

- **Ontology-based information extraction works better, but**
 - **is it specific enough for TPS?**
 - **will it carry over to other application domains?**

Challenges with Cognitive Services

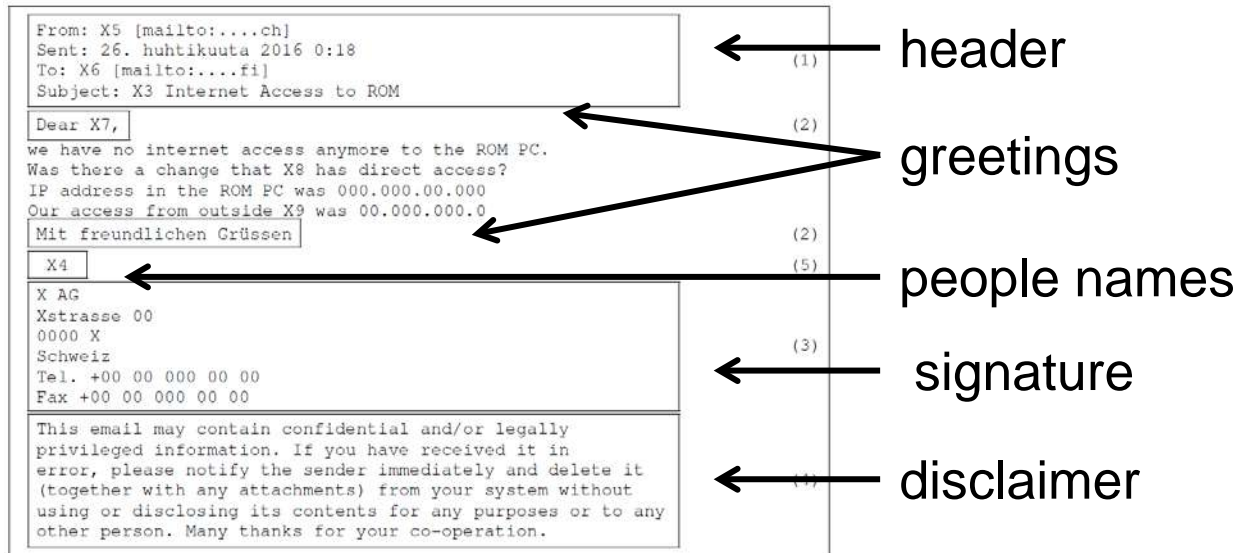


The Extractor Service



- de Carvalho/Cohen: Learning to extract signature and reply lines from email [CEAS, 2004]
 - Sequence of lines approach
 - Pattern-based feature vector for each line
 - Used to train an ML classifier
- + Matching of multi-language patterns
- + Arbitrary conversations

Pattern-based Textline Features



*in arbitrary order
and multiple
languages*

pattern	regular expression
salutation (g)	$\backslash s \backslash \backslash A ((S s)ehr\ gee?hrter? (H h)err\ (en)? (F f)rau (D d)amen Liebe\ (r s) (L l)ieber? (L l)iebster? (H h)allo\ (H h)ey G g)rezi (H h)\ (I i) (M m)orgen? (G g)ue?ten?((Tag Abend Morgen Obig) (H h)oi),?.?(\s \Z)$
greeting (g)	$\backslash s \backslash \backslash A ((G g)r\ (ue u)\ (ss e)n? (A l)les\ ((G g)ute (G g)ueti) (H h)erzlich Alles\ (L l)ieb(e i) D d)anke? (S s)ali (V v)ielen?),?.?(\s \Z)$
salutation (e)	$\backslash s \backslash \backslash A ((D d)ear (D d)earest (H h)ello (H h)ey (M m)y\ (D d)ear),?.?(\s \Z)$
greeting (e)	$\backslash s \backslash \backslash A ((R r)egards? (T t)hanks (C c)heers (G g)rateful (S s)incerely),?.?(\s \Z)$

Number of Patterns

10 greetings

16 headers

113 signatures

Threshold-based Classifier vs. Trained Classifiers

Category	No. of lines	FP	FN	TP	TN	Accuracy (%)	$F_{0.5}$ (%)
header	2065	13	198	1867	24210	99.19	97.39
greeting	1853	150	523	1330	24285	97.44	85.55
signature	8120	3696	2028	6092	14472	78.22	64.44
disclaimer	1167	199	868	299	24922	95.94	47.33
name	585	987	254	331	24716	95.28	28.26
problem description	12498	2551	3725	8773	11239	76.13	75.90

Feature Vector	Learner	FP	FN	TP	TN	Accuracy (%)	$F_{0.5}$ (%)
signature only	SVM	237	588	1052	3381	84.31	77.40
signature only	AdaBoost	216	613	1027	3402	84.23	77.66
signature only	RF	142	489	1151	3476	87.99	84.48
sign. + header	SVM	220	573	1067	3398	84.91	78.59
sign. + header	AdaBoost	216	552	1088	3402	85.39	79.35
sign. + header	RF	155	444	1196	3463	88.61	84.90
all features	SVM	215	553	1087	3403	85.39	79.37
all features	AdaBoost	219	527	1113	3399	85.81	79.87
all features	RF	148	422	1218	3470	89.16	85.73

Outlook

- Pattern-based feature generation and ML-based approach will be further refined for Extractor Service
- Semantic Analysis needs to be strengthened
 - Experiments with IBM NLU will continue
 - Own ideas following an ontology-based information extraction approach
 - Easy transfer to other application domains
- TPS similarity based on
 - *Andreas Ecke: Quantitative methods for similarity in description logics, Dissertation, TU Dresden, 2016*
- Skill map 0.9 go live end of 2017