Common Sense.
For Computers.
Artificial Intelligence in Requirements Engineering
Dr. Mathias Landhäußer
Dr. Sven J. Körner
95% of Requirements are Recorded in Natural Language

Getting the facts right is not enough!
Challenge: from Requirements to Software

- Requirements (Stakeholder)
- Models (Analyst)
- Software (Development)

Missing Expectations?

Erosion

Mostly Manual Processes
AI domain of expertise is very limited to whatever universe we train them on.

Most of the systems, you show them [...] unusual situations [...] and they will say complete garbage about it.

They don't have common sense.

Yann LeCun, Facebook AI
What Is Common Sense?

• The trophy does not fit into the suitcase, because it is too big.

• The trophy does not fit into the suitcase, because it is too small.

Why do you know, what “it” refers to?
“People remember errors committed by AI, but forget human errors”
Linguistic Flaws in Requirements

Linguistic Flaws

Generalization
- Presuppositions
- Incomplete comparatives and superlatives
- Modal words (possibilities)
- Modal words (necessities)
- Incompletely specified process words

Distortion
- Quantifiers
- Incompletely specified conditions
- Nouns without reference index

Deletion
- Nominalizations
No Tools, Just Rules?

http://www.sxc.hu/photo/1269809
43% of all errors in IT and engineering projects lead back to wrong specifications.

Today, errors based on meaning and understanding must be solved by humans.

Comprises 95% of all specifications. Also, natural language is the means of choice for anybody to communicate with computer systems.

Are a key aspect to cognitive computing challenges which cannot be solved with machine learning (neural networks) and statistical methods.

80% of data today is „dark“. By 2020, 93% of data will be „dark“.
Replacing Drudge Work
RESI: The Technical Approach

Specification Text → Model

NLP → Ontologies

Apply Rules

Model → Improved Model

Export Model to Text
Semantic Processing

The Semantic Model is an annotated parse tree, enriched with thematic/semantic role labeling and further semantic information to semantic concepts.

The information layer comprises ontologies, knowledge graphs, lexicons, statistics, and NN to challenge semantics from above deducted model.

The decision layer uses “common sense” to make meaning of the semantic model and augments it. This rule processing is an n-tier approach to solving semantic queries.

Semantic Rule collections to augment:
- Bots (Virtual Assistants)
- LegalTech
- Requirements Engineering
- FinTech + Tax + Auditing
- InsurTech
- RetailTech

Think Humanly
Ontologies offer world knowledge to a computer system. They provide semantics and therefore the meaning of a sentence.
RESI Integrated into ProContext’s ProceManManager

**Edit requirement "UR-35"**

* Type: Requirement for use

**Example:** The user must be able to recognise by the system at first sight which order must be worked on as the next.

If an appointment is cancelled, the system must display a message.

- **Unclear determiners**
  Please specify a determiner for:
  1: "an" is an unclear determiner.
  3: "the" is an unclear determiner.
  6: "a" is an unclear determiner.

- **Unspecified process words**
  There seem to be details missing for the following process words:
  2: "cancelled" seems to be incompletely specified.
    Missing arguments: Performed by, Evallee-Direct, Purpose in event
  5: "display" seems to be incompletely specified.
    Missing arguments: Sender of info
    Detected arguments: Instrument-Generic (system)
Even Non-Professionals Can Improve Specs!

Flaws Identified Manually vs. Automatically in Monitoring Pressure Text

- Group N
- Group P
- Group D

Monitoring Pressure RESI  Monitoring Pressure manuell
if it’s not working

it better be the customer’s fault
You can observe a lot by watching.

Yogi Berra
Threats to Validity / Issues / Problems

• Internal Validity: case studies in research show the validity of the approach in known use-case scenarios and specifications

• External validity: first results come from demonstrators, but we need to gather more data to being able to make a real statement

• No answer to the question: When can we ignore flaws, when are they important?

Integrating into everyday workflows (IBM Doors, Jira, PTC, Polarion)

• Biggest problem:
  • finding real-life requirements
  • finding companies that are willing to share their experience in RE openly
IN THEORY, THEORY AND PRACTICE ARE THE SAME. IN PRACTICE, THEY ARE NOT

— Albert Einstein —
我看到飞机飞行。
Wǒ kàn dào fēijī fēixíng.
我看到飞机飞行。
I saw the plane flying.
How Google et al. Work
I saw the plane flying.
How Google et al. Work

I saw the **plane** flying.

I saw the **mountains** flying.
Three Main Approaches to AI

**Statistics**
- Better for non-complex relationships in data
- Can rate results with confidence
- Deals with uncertainties
- Fast for not-so-complicated systems
- Expensive training
- Parametric model requires statistical knowledge
- Error prone in parameter estimation

**Machine Learning / Deep Learning**
- Ability to detect complex nonlinear relationships between dependent and independent variables
- Works great for perception already today
- Easily implemented (i.e. in multicore processors or systems with GPUs)
- Needs Supervised Learning (which limits the machine power through mankind)
- Does not work with low sample size
- Black box (rather difficult to interpret and to explain/to rebuild)
- Retraining is hard (retraining for backpropagation is problematic)
- Can’t do a priori

**Semantics**
- Understands the meaning of natural language
- Complements statistical and ML approaches
- Can justify
- Works a priori
- Needs (linguistic) experience
- Computing power
- Quality depends on ontology (semantic knowledge database)
- Not a one-stop shop (complements other approaches)
A Little Brain Teaser

Killing
BAD

Killing Bacteria
GOOD

Failing to Kill Bacteria
BAD

Never Failing to Kill Bacteria
GOOD

Understanding the meaning of text continues to require knowledge of who produced it and who it is aimed at.
DeNom

Special Treatment for Nominalizations
Nominalizations: Problematic yet often overlooked

• Nominalizations can lead to serious problems during development

• A requirements engineer’s writing rule: Though shall not use nominalizations!

• Inspection rule: Find and eliminate all nominalizations!
  • Can be identified automatically using RESI [RESI]
  • RESI is picky and produces many warnings
  • Effort too high for real-world scenarios [RESI@Automotive]
Not All Nominalizations are Problematic

Linguistic Flaws

- Generalization
- Distortion
- Deletion

Nominalizations

Category 1: self-descriptive
Category 2: defined in the sentence-wide context
Category 3: defined in the document-wide context
Category 4: underspecified
Fun Fact: Most Nominalizations are OK!

Fully Manual Study:
- 5 specifications
- >40,000 words
- 356 nominalizations in total

- 0 % Category 1
- 70 % Category 2
- 29 % Category 3
- 1 % Category 4

Half-automated Study:
- 6 specifications
- >33,000 words
- 499 nominalizations detected

- 0 % Category 1
- 83 % Category 2
- 8 % Category 3
- 0.2 % Category 4
+ some false positives
Automatic Categorization

Nominalizations identified by RESI

Glossary (List of Words)

Is the nominalization part of a nominal phrase?

Consider sentence context (see next slide)

No Context Detected

Self-descriptive → Category 1

Defined in the sentence-wide context → Category 2

Category 3 or 4 → Warning
Evaluation

- 10 specifications, >59,000 words
- 1,136 nominalizations
  - only 84 of them are problematic
  - DeNom shows 129 warnings
- Precision of RESI on average: 8% ($F_1=15\%$)
- Precision of DeNom on average: 65% (with a recall of 88%, $F_1=75\%$)
Product: Interactively Disambiguate Requirements Specifications
References


Contact Us

thingsTHINKING GmbH
Eisenhutstraße 14
76703 Kraichtal, Germany

Dr. Mathias Landhäußer
+49 175 1970 705
mathias@thingsTHINKING.net

www.thingsTHINKING.net