

Assisted Requirements Engineering

Or: What will remain in the hands
of the future requirements engineer?

SWQD 2022

Prof. Dr. Henning Femmer

 Qualicen



So, what does RE today look like?

In Automotive...

Screenshot of the DOORS Requirements Management tool interface showing a test report view.

The interface includes a menu bar with File, Edit, View, Insert, Link, Analysis, Table, Tools, Discussions, User, Publish, RQM, Change Management, Help, and a toolbar with various icons.

The main window displays a table titled "RQM Test Report View" with columns for ID, Design Input, Test Cases, Test Status, Verdict, Passed, Deferred, Incomplete, Inconclusive, Error, Partially Blocked, Blocked, Failed, Perm Failed, and Verdict Changed.

Key requirements listed include:

- PM_821**: **1 Measurement algorithms are accurate, reliable, easy to calibrate and age appropriate**
 - Provide lead selection same as bedside
 - Provide 6 - 8 sec. ECG wave + cascade (freeze) capability
 - Full channel flexibility including the 2 V leads.
- PM_822**: Provide lead selection same as bedside
- PM_823**: Full channel flexibility including the 2 V leads.
- PM_824**: Achieve superior clinical contributions in these measurements. Integrate market leader technologies when appropriate.
- PM_825**: Same algorithm but limited alarms and event storage Vfib, Vtach, Vrungs, PVCs, Asystole, Brady, Pacer, pro-pacer not pacing or pacer not captured by heart.
- PM_826**: Graphical and tabular trends, selected alarms and arrhythmias stored. Events easily marked. Timing appropriate for application area.
- PM_827**: **2 Tailored monitoring for diverse set of patients and improved method to admit, se**
 - Events need to be captured for later review.
 - Life threatening/arrhythmia alarms on instantaneously
 - Permit alarms to go below preset factory defaults (resp)
 - Clearer method which is intuitively obvious to the user
 - Ensure message clarity and user understanding
 - Includes electronic and printed records
- PM_828**: Life threatening/arrhythmia alarms on instantaneously
- PM_829**: Permit alarms to go below preset factory defaults (resp)
- PM_830**: Clearer method which is intuitively obvious to the user
- PM_831**: Ensure message clarity and user understanding
- PM_832**: Includes electronic and printed records
- PM_833**: **3 Measurement algorithms are accurate, reliable, easy to calibrate and age appropriate**
 - Provide lead selection same as bedside
 - Provide 6-8 sec. ECG wave + cascade (freeze) capability
 - Full channel flexibility including the 2 V leads.
- PM_834**: Provide lead selection same as bedside
- PM_835**: Provide 6-8 sec. ECG wave + cascade (freeze) capability
- PM_836**: Full channel flexibility including the 2 V leads.

For today's cars:
Several million
requirements...

For Business Information Systems

Use Case 2 Get Paid for Car Accident

Primary Actor: Claimant

Scope: Insurance company ("MyInsCo")

Level: Summary

Stakeholders and Interests:

Claimant—to get paid the most possible.

MyInsCo—to pay the smallest appropriate amount.

Department of Insurance—to see that all guidelines are followed.

Precondition: None.

Minimal Guarantees: MyInsCo logs the claim and all activities.

Success Guarantees: Claimant and MyInsCo agree on amount to be paid; claimant gets paid that.

Trigger: Claimant submits a claim.

Main Success Scenario:

1. Claimant submits claim with substantiating data.
2. Insurance company verifies claimant owns a valid policy.
3. Insurance company assigns agent to examine case.
4. Insurance company verifies all details are within policy guidelines.
5. Insurance company pays claimant and closes file.

Extensions:

1a. Submitted data is incomplete:

- 1a1. Insurance company requests missing information.
- 1a2. Claimant supplies missing information.

2a. Claimant does not own a valid policy:

- 2a1. Insurance company denies claim, notifies claimant, records all this, terminates proceedings.

3a. No agents are available at this time.

- 3a1. (What does the insurance company do here?)

4a. Accident violates basic policy guidelines:

- 4a1. Insurance company denies claim, notifies claimant, records all this, terminates proceedings.

4b. Accident violates some minor policy guidelines:

- 4b1. Insurance company begins negotiation with claimant as to amount of payment to be made.

For a global notification and license renewal software:

> 600

use cases.

The „modern“ way

168 Search by Name

As a help desk operator I want to search for my customers by their first and last names so that customer response times remain short

For a simple webshop:
>1000
user stories.

What a mess.

What a mess.
Can't we "AI" this?



WELL'S, FARGO & COMPANY

Special Exhibit –
The Butter

Since the beginning of 1852,
Wells Fargo & Company
has provided safe and
reliable banking services.

Imagine the excitement
of traveling across the country
in a Wells Fargo stagecoach.
Deliveries twice a week, Wells

Fargo arrived in San

Francisco from St. Louis

in just 23 days.

Wells Fargo is proud to

offer a transcontinental stagecoach

The Overland Mail.



In 1852,
Wells Fargo & Company
began its first stagecoach
service between San Francisco
and St. Louis.

They have
continued to
provide reliable
banking services

and travel services

ever since.

Experience the

stagecoach.

All aboard the



Concord Coach No. 251
La Diligencia Concord N° 251

Please Do Not
Touch
Por Favor No
Tocar

An Overland Journey
Un Viaje por Tierra

Para los Jóvenes en las grandes
caminatas.

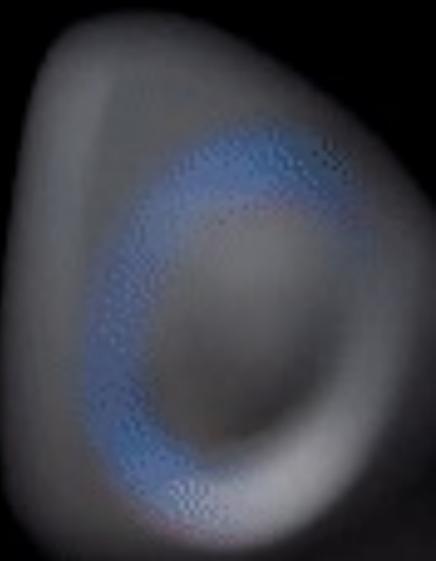
Conjunto para los Jóvenes en las grandes
caminatas.

As a car driver,
I want to change the direction of the car,
so that I can reach my desired destination.









AutoX Opens Its Fully Driverless RoboTaxi Service to the Public in China (English)





Phase 4: Fully automated.

Humans only define their goal, the right means as well as execution of those is performed by computers.



Phase 3: Partly automated.

Systems take over simple tasks and actively prevent common errors.



Phase 2: Assistance Function.

Systems provide guidance and passively prevent common errors.



Phase 1: Manual work.

Manual choice of means,
indirect link to physical world.



As a car driver,
I want to change the direction of the car,
so that I can reach my desired destination.



As a car driver,
I want to ~~change the direction of the car,~~
~~so that I can~~ reach my desired destination
automatically.

Revolutionary is **not** if you deliver on what people **ask** you to build,
Revolutionary is if you deliver on the why.

As a requirements engineer,
I want to write requirements,
so that we have a complete and consistent system specification.
Can we “AI” this?

As a requirements engineer,
I want to ~~write requirements,~~
~~so that we~~ have a complete and consistent system specification.
Can we "**AI**" this?

As a requirements engineer,
I want to write requirements,
so that we have a complete and consistent system specification.
Can we automate this?

Or: Verify the product.

Or: receive clear instructions for our developers.

Or:

...

Or:

...

Or:

...

Automation in Requirements Engineering

?



Phase 2: Assistance Function.



IBM Rational
DOORS Next Generation



Phase 1: Manual Work.





As far as possible, the system should check the input for plausibility.



Bad requirements – We've all seen them and we all know about their risk.
But what can you do, right?



Takes **>4** weeks.

30k-50k€ per Review

“Do you remember our guidelines? Really? You must be the only one.”

Manual requirements reviews are **costly, inconsistent and slow.**

But what can you do, right?

Show items containing where

S...	Text	QRC_Findings	QRC_Findings_Text
	Old:	New:	
6	The vehicle is used in a range between 4 to 15 km/h during normal operation.	<input type="checkbox"/> Passive Voice	The vehicle is used in a range between 4 to 15 km/h during normal operation.
7	The document handler can refine the search results and set further filters by using the facets provided on the left side of the screen.	<input type="checkbox"/> Vague Words <input type="checkbox"/> User Interface Details <input type="checkbox"/> Long Sentences	The document handler can refine the search results and set further filters by using the facets provided on the left side of the screen.
8			

Qualicen Holmes.
Fast, objective feedback right when and where you need it.

What can automatic methods do for you?

Helping from Two Perspectives

The diagram illustrates two perspectives on requirements analysis, connected by orange arrows pointing from the central text boxes to their respective interface screenshots.

Authors (Orange Side):

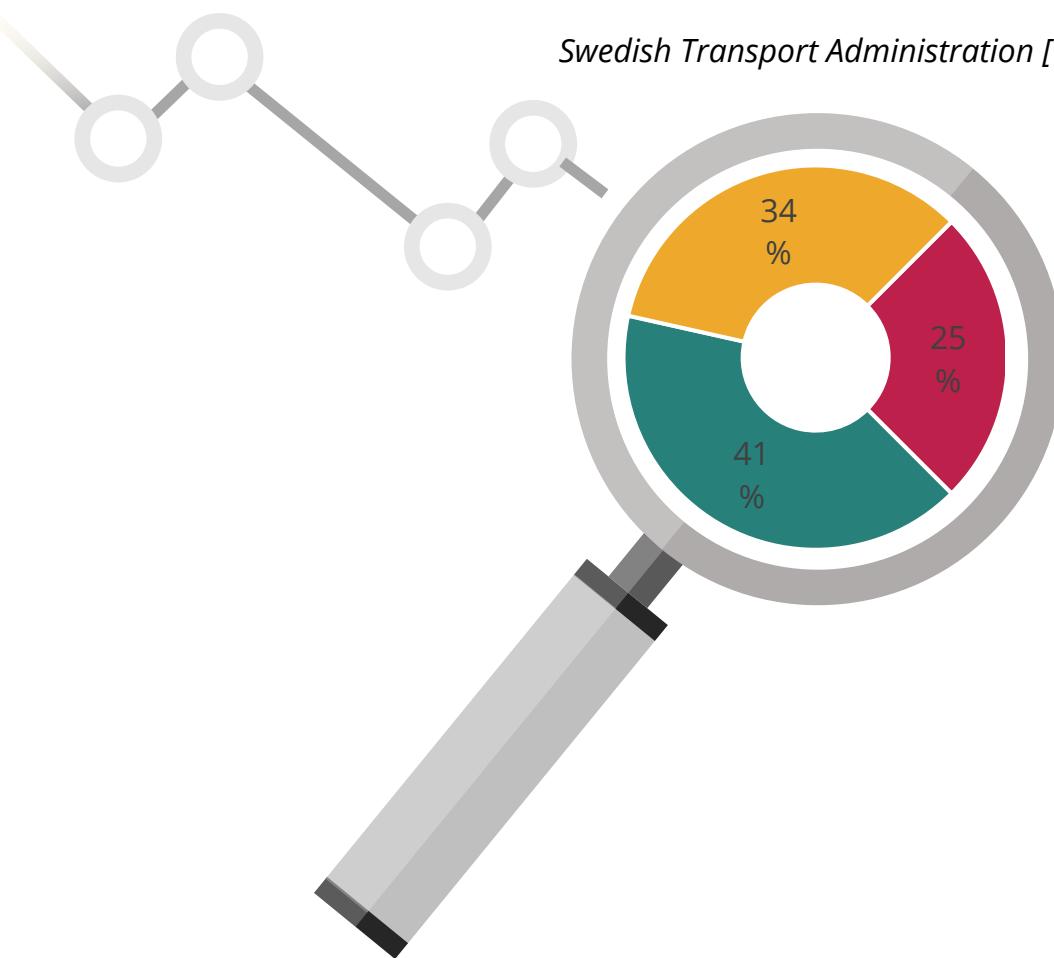
- Immediate feedback:** Get direct feedback regarding quality problems in your requirements. (Screenshot shows a comparison between old and new requirement text with red highlights and error markers.)
- Update on save:** The problem markers are updated as soon as you save the document. (Screenshot shows a comparison between old and new requirement text with red highlights and error markers.)

Quality Managers (Teal Side):

- Analysis over time:** Check whether your quality measures work and get an early warning if something goes astray. (Screenshot shows a dashboard with KPIs like Current Size in Words, Current Number of Files, and Current Number of Errors, along with a metrics trend chart and a treemap for finding density.)
- Hotspot localization:** Find out where findings accumulate. This helps you improve effectively. (Screenshot shows a metrics trend chart for 'qualicenscout' with a green dot indicating a peak in findings in October.)
- Query findings:** Findings are saved with the requirements. You can use them for queries or reports. (Screenshot shows a commit chart for 'qualicenscout' with a green dot indicating a peak in commits in April.)
- Key Performance Indicators:** Customizable KPIs are visible on the dashboard and updated with every save action. (Screenshot shows a dashboard with various KPIs and a 'no change' status message.)

What can automatic methods not do for you?

Results from a systematic analysis.



■ Deterministic

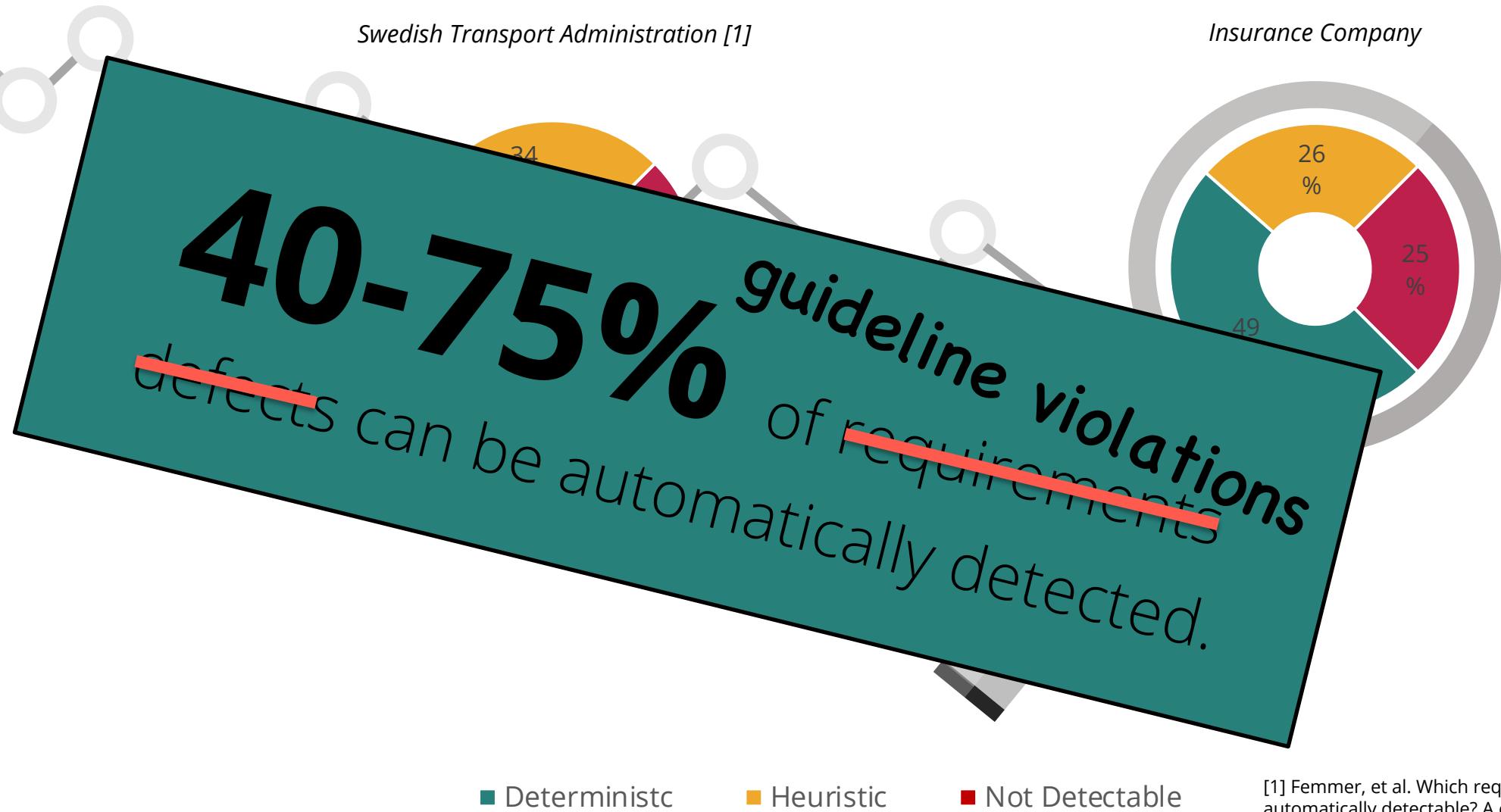
■ Heuristic

■ Not Detectable

[1] Femmer, et al. Which requirements artifact quality defects are automatically detectable? A case study. AIRE, 2017

What can automatic methods not do for you?

Results from a systematic analysis.



[1] Femmer, et al. Which requirements artifact quality defects are automatically detectable? A case study. AIRE, 2017

What can automatic methods do for you?

Some of the many built-in checks

The screenshot shows the Scout application interface with the 'Projects' tab selected. The interface is divided into four main sections: Lexical Smells, Grammatical Smells, Structural Smells, and Consistency.

- Lexical Smells:**
 - Imprecise Phrases
 - External Links
 - Text in Brackets
 - Dangerous Slashes
 - Universal Quantifiers
 - Spelled Out Numbers
 - Negative Words
 - UI Details
 - Occurrence of 'will' or 'may'
 - Triple Dots
 - Trivial Shall Be
- Grammatical Smells:**
 - Passive Voice
 - Grammar Mistake
 - Spelling Mistake
 - Vague Pronouns
 - Comparatives
 - Superlatives
 - Improper Conditions
 - Long Sentences
- Structural Smells:**
 - Exactly One Shall or Should
 - Default Texts
 - Long requirement
 - Example
 - EARS Ubiquity
 - EARS Non-Conformance
 - Empty section
- Consistency:**
 - Non SI Unit
 - ToDos
 - Usage of Synonyms
 - Abbreviations



The Bleeding Edge: Semantic Requirements Clones Detection at Scale

Project: De... Branch: default-branch Timetravel: Current Documents

Actions ▾

demo / [Integrity] / Demo (2634) / 2659

Wenn der Auftragnehmer sich widersprechende Anforderungen feststellt, dann muss der Auftragnehmer dieses dem Auftraggeber sofort ⚠ nach Kenntnisnahme in der dokumentierter Form mitteilen. ⚠

Basic Tool: Semantic Similarity

When the address is successfully validated by the case handler, a confirmation mail is sent out by the system.

$$\begin{pmatrix} 0.14 \\ 0.47 \\ \dots \\ 0.94 \end{pmatrix}$$

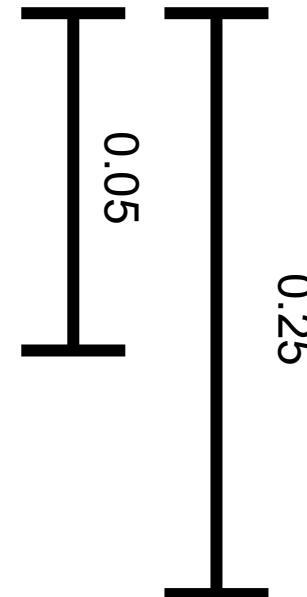
When the state of the address is set to 'verified', the system generates a confirmation message.

$$\begin{pmatrix} 0.14 \\ 0.37 \\ \dots \\ 0.84 \end{pmatrix}$$

The sky is blue, the tree is green and the sun is shining.

$$\begin{pmatrix} 0.5 \\ 0.61 \\ \dots \\ 0.34 \end{pmatrix}$$

STEP 1: Vectorization
(e.g. n-grams,
word2vec, BERT, etc.)



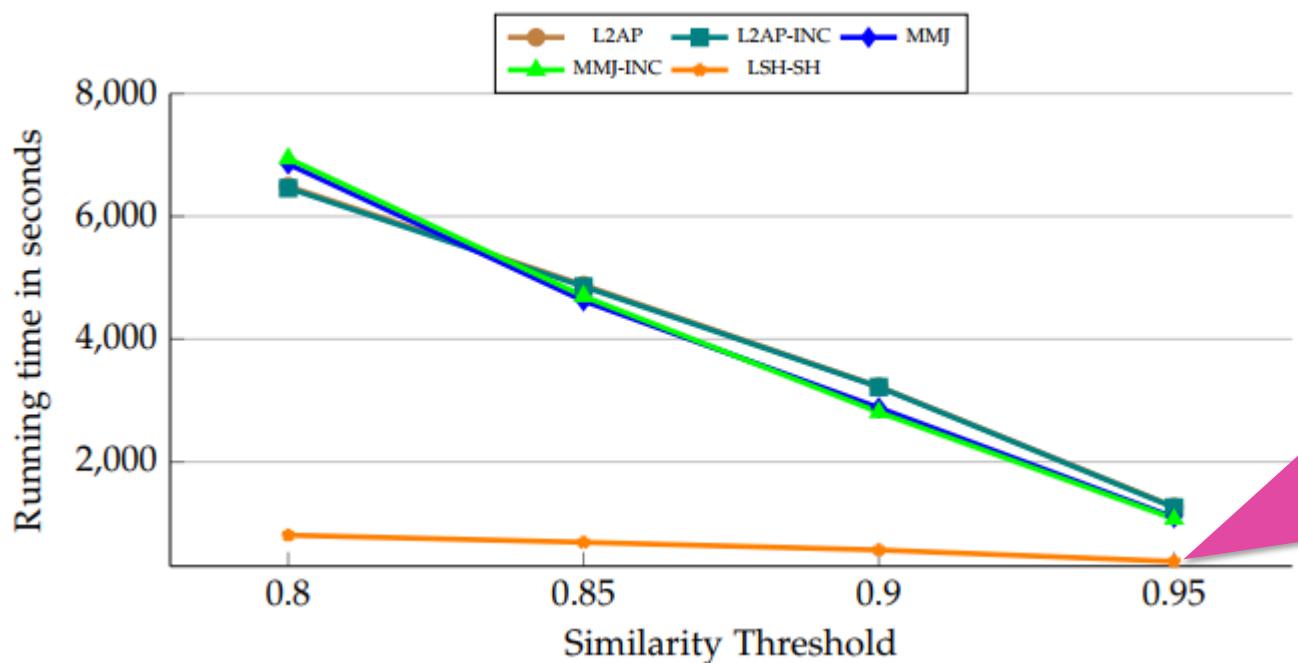
STEP 2: Calculate Distance



STEP 3: Flag Clones

Challenge: Semantic Clone Detection at Scale

Analysis time for 200.000 Requirements



Lukas Konwitschny, 2020

Depending on chosen semantic representation and similarity threshold, finding all clone pairs can be as fast as

**~2,5 minutes
for
200.000 requirements
(20 billion potential pairs)**

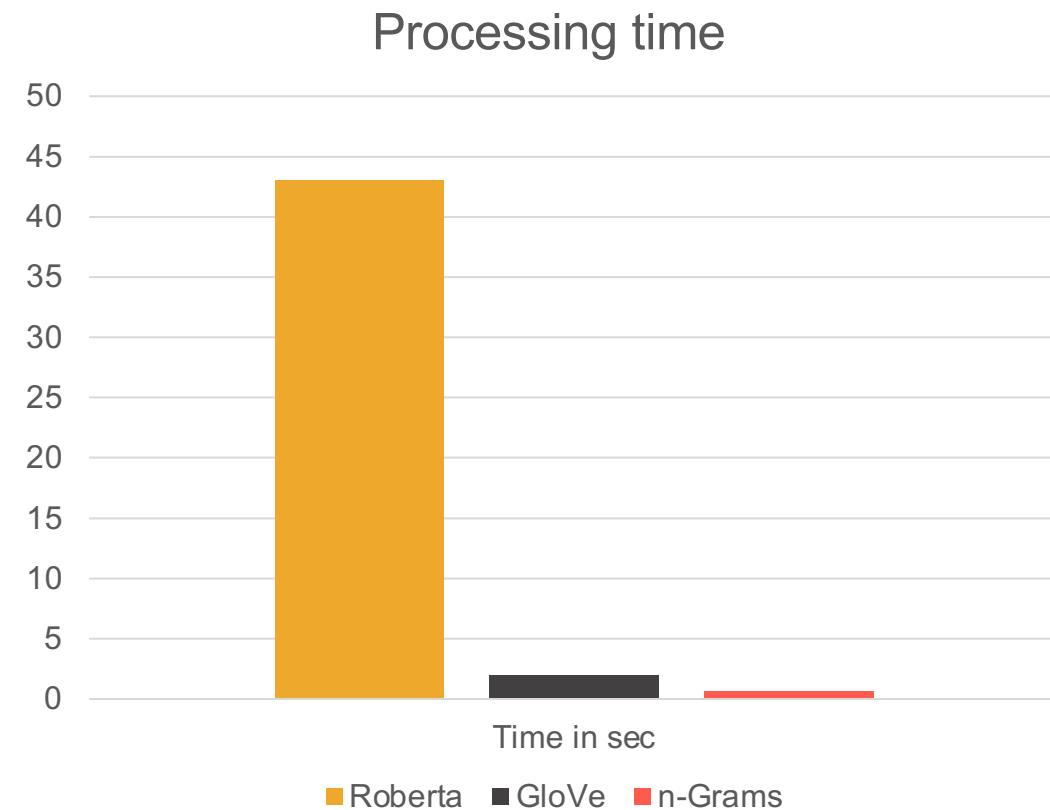
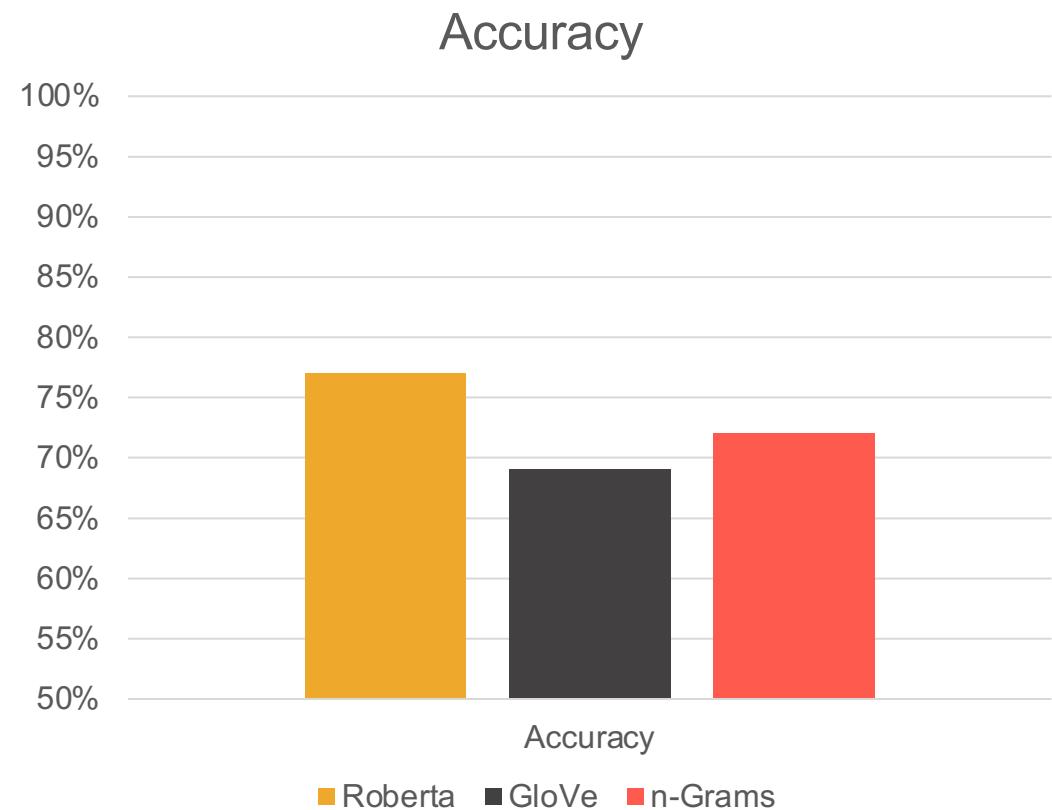
Experiment: Comparison of Models for Semantic Similarity

1. Roberta-base Transformer Models (125 Mio Parameters)
Fine-tuned for semantic similarity version from s-bert
2. GloVe Word Vectors: Pre-Calculated Vectors for single words.
Averaged to obtain sentence vectors (via s-bert)
3. n-Grams: Count which n characters occur together
„This is a test“ → Th:1, hi:1, is:2, ... → (1,1,2,...)

Conditions:

Fixed test set of sentence pairs
Executed on cloud with V100 GPU

Comparison of Models for Semantic Similarity



Automation in Requirements Engineering

?



Phase 3: Partly automated.

Smell Analysis



Phase 2: Assistance Function.



IBM Rational
DOORS Next Generation

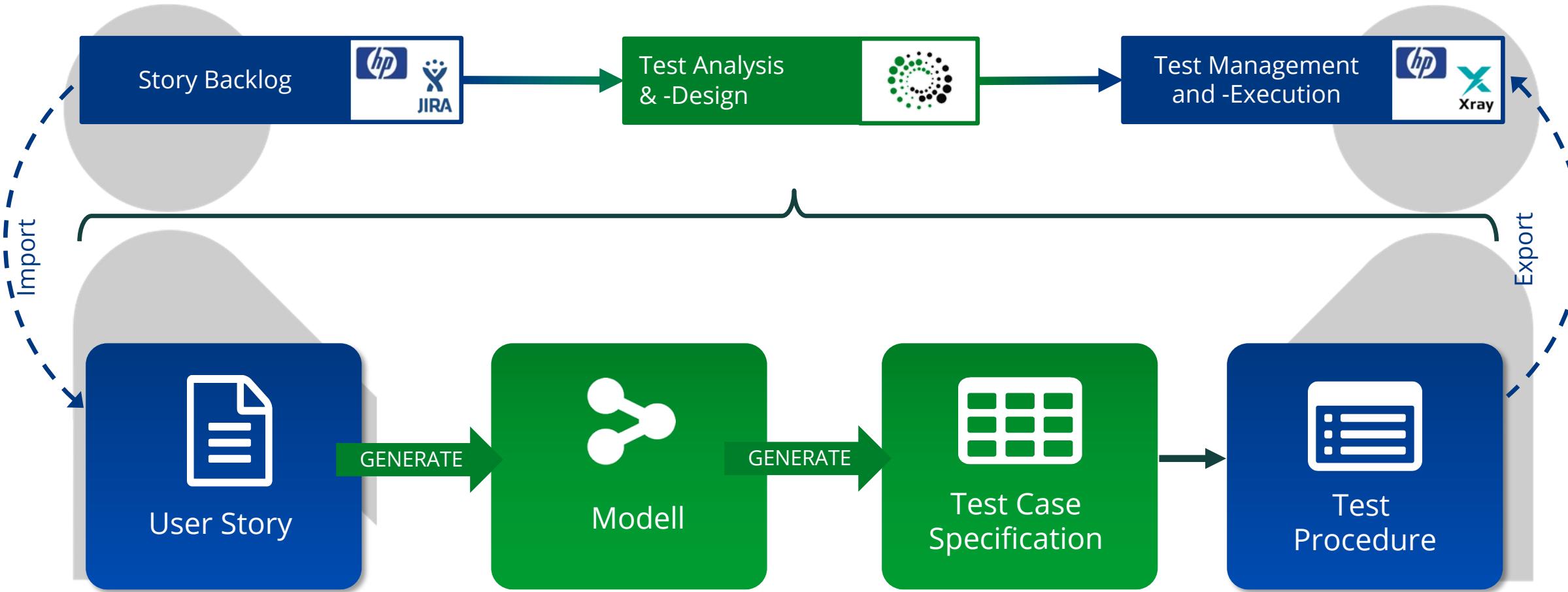


Phase 1: Manual Work.



As a requirements engineer,
I want to ~~create a test case specification,~~
~~so that I~~ know whether my requirements are fulfilled.
Can we "AI" this?

Test Analysis and -Design with Specmate



Model & Test-Generation in Action

Causality in Requirement Artifacts (CiRA)



[Home](#) [Demo](#) [Sessions](#)

Causality Extraction: Demo

Automatic, step-wise extraction of cause-effect-relations from natural language sentences.



Step 1: Classification

As a first step it needs to be determined whether the sentence is causal at all. Enter a sentence below and execute the binary classifier.

If the user doesn't have a login, and login is needed, or an error is detected, a warning window is shown and a signal is emitted.

Classify

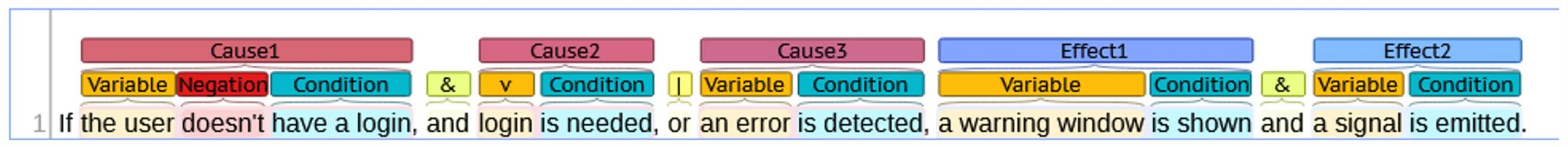
<http://cira.diptsrv003.bth.se/extractor/classify>

Model & Test-Generation in Action

Step 2: Labeling

In the second step the causal sentence has to be dissected: a labeling algorithm annotates each word depending on its contribution.

The labeling process may take a few seconds before being displayed.

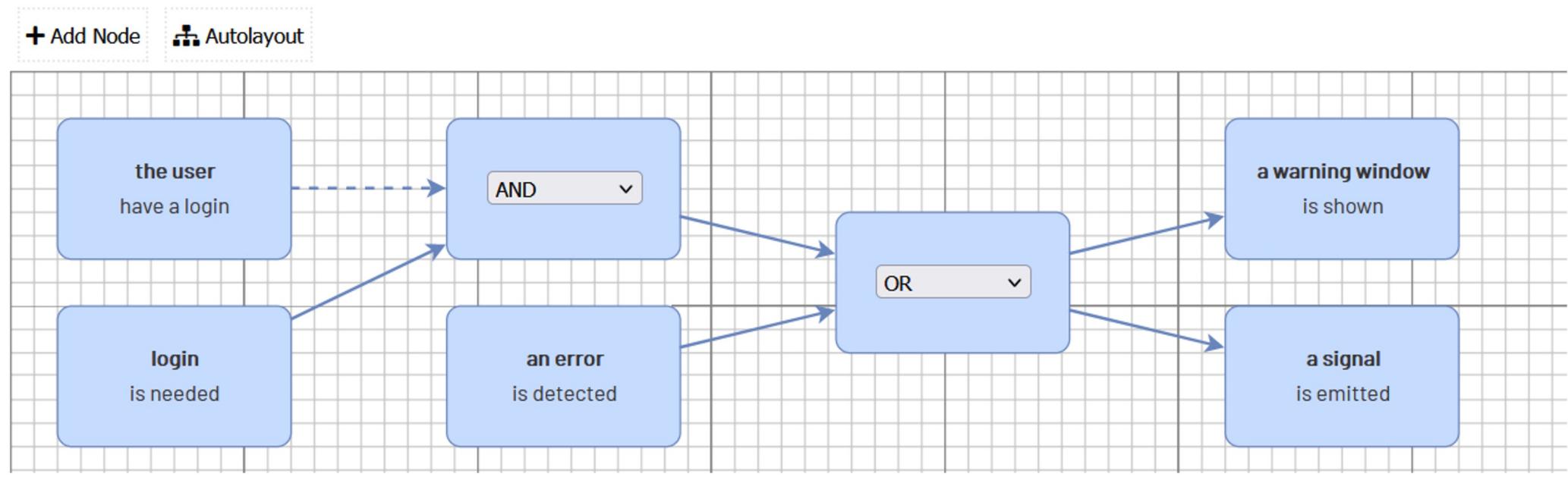


Model & Test-Generation in Action

Step 3: Graph Generation

In the third step of the process the labeled sentence is transformed into a Cause-Effect-Graph (CEG). If you disagree with the automatically generated CEG you can adapt the graph at this point.

If the user doesn't have a login, and login is needed, or an error is detected, a warning window is shown and a signal is emitted.



Model & Test-Generation in Action

Step 4: Test Case Generation

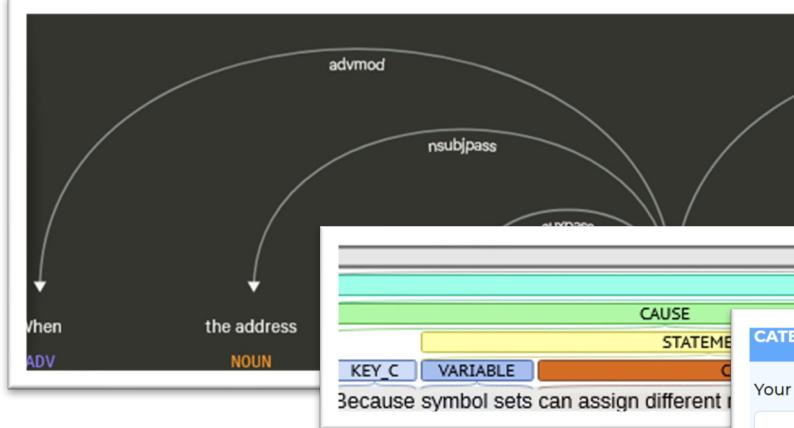
Finally, the graph is transformed into a test suite with a minimal amount of test cases.

If the user doesn't have a login, and login is needed, or an error is detected, a warning window is shown and a signal is emitted.

ID	Input			Expected Result		E
	the user	login	an error	a warning window	a signal	
1	not have a login	is needed	not is detected	is shown	is emitted	
2	have a login	is needed	is detected	is shown	is emitted	
3	not have a login	not is needed	is detected	is shown	is emitted	
4	have a login	is needed	not is detected	not is shown	not is emitted	
5	not have a login	not is needed	not is detected	not is shown	not is emitted	



Model & Test-Generation in Action



```
def rule Condition4_3 {  
    [Effect] - dobj ->  
    [Effect_SubA] - prep ->  
    IN: 'as' - pobj ->
```

First Attempt: Syntactical and Grammar Rules

More and more rules needed
→ rules did not generalize well

Manually label ~2.000 texts

Train Tree-Recursive Neural Networks to learn the hierarchical structure.

For „complex“ examples, accuracy <2%

<https://causalitytreeextractor.com/>

CATE: CAusality Tree Extractor from Natural Language Requirements

Your Sentence

When the system crashes, a warning message is shown.

Beam Width: 10

Temperature Scaling: BERT/Left Branching only

Dataset: Left Branching, Right Branching

Word Embeddings: BERT, FastText, Random (300)

Predict

Got result in 1.65s

GoJS 2.1 evaluation
(c) 1998-2021 Northwoods Software
Not for distribution or production use
gojs.net

No Co-Reference Resolution, no reformulation active/passive etc.

Let's see what GPT-3 makes of this*

Idea: Translate texts to a semi-formal representation that can be easily parsed automatically

Input: Tarifs that are not applicable for a grant are hidden on the letter, unless they are marked as applicable.

Output: IF “Tarifs are not applicable for a grant” AND NOT “Tarifs are marked as applicable”
THEN “Tarifs are hidden on the letter”

* This was in a few-shot setting with 10 examples, none of which resembled this example

Current Approach

Intermediate Results:
Accuracy > 70%

Token Classification 2
(using BERT Transformer)

⋮
⋮
VAR CONDITION OR VAR CONDITION
If the user enters a wrong input or the system fails to open
the login mask, a warning window shall be opened.
VAR CONDITION

Token Classification 1
(using BERT Transformer)

CAUSE OR CAUSE
If the user enters a wrong input or the system fails to open
the login mask, a warning window shall be opened.
EFFECT

Automation in Requirements Engineering



Phase 4: Fully automated.

Creating SE artifacts



Phase 3: Partly automated.

Smell Analysis



Phase 2: Assistance Function.



IBM Rational
DOORS Next Generation

MODERN Requirements

ReQtest

jama software®

osseono



Phase 1: Manual Work.

Accompa

MICRO FOCUS

pearls
The Solution Is Simple

HelixALM

ViSURE

Some Inspiration from Software Development

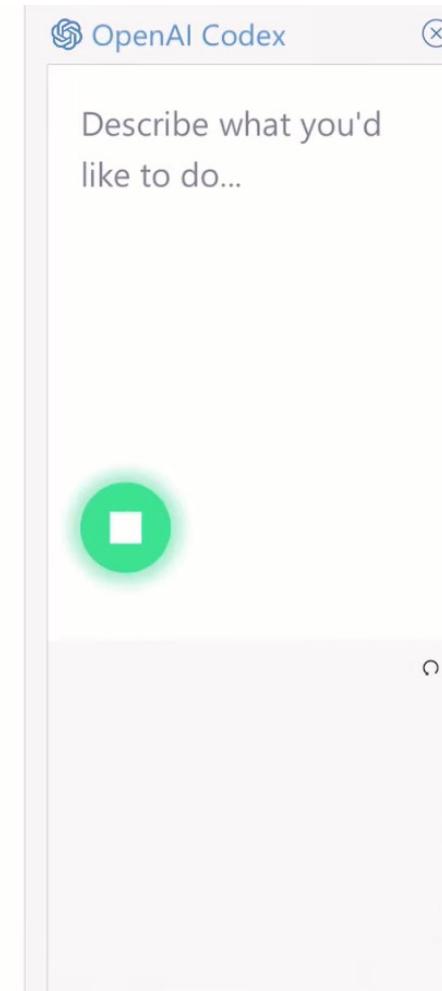
'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

“Beware the Jabberwock, my son!
The jaws that bite, the claws that catch!
Beware the Jubjub bird, and shun
The frumious Bandersnatch!”

He took his vorpal sword in hand;
Long time the manxome foe he sought—
So rested he by the Tumtum tree
And stood awhile in thought.

And, as in uffish thought he stood,
The Jabberwock, with eyes of flame,
Came whiffling through the tulgey wood,
And burbled as it came!

OpenAI Codex API Documentation



Custom prompt

Type something and a neural network will guess what comes next.

COMPLETE TEXT

What will stay in the hand of future REs? – And what not?

Future REs will concentrate on

- Setting the scope
- Define what is needed
- Performing social activities
- Setting priorities, resolving conflicts, finding consensus

(AI) Assistants will transform the input

- (Re-) formulate requirements given an input
- Ask-/Answer questions





Let's make
this real.

prof dr henning femmer

+49 176 38027921

henning.femmer@qualicen.de | femmer.henning@fh-swf.de

 **Qualicen**
great methods and tools
for every engineer.

Fachhochschule
Südwestfalen 
University of Applied Sciences

Most Relevant Publications

- H. Femmer. "Assisted Requirements Engineering-What Will Remain in the Hands of the Future Requirements Engineer?" International Conference on Software Quality. Springer, Cham, 2021.
- H. Femmer and A. Vogelsang, "Requirements Quality Is Quality in Use," *IEEE Softw.*, vol. 36, no. 3, 2019.
- H. Femmer, D. M. Fernández, S. Wagner, and S. Eder, "Rapid quality assurance with Requirements Smells," *J. Syst. Softw.*, vol. 123, 2017.
- H. Femmer, M. Unterkalmsteiner, and T. Gorschek, "Which Requirements Artifact Quality Defects are Automatically Detectable? A Case Study," in IEEE 25th International Requirements Engineering Conference Workshops, AIRE, 2017.
- J. Fischbach, J. Frattini, D. Mendez, M. Unterkalmsteiner, H. Femmer, & A. Vogelsang. How Do Practitioners Interpret Conditionals in Requirements?. In *International Conference on Product-Focused Software Process Improvement* (pp. 85-102). Springer, Cham, 2021.
- J. Fischbach, T. Springer, J. Frattini, H. Femmer, A. Vogelsang & D. Mendez. Fine-grained causality extraction from natural language requirements using recursive neural tensor networks. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)* (pp. 60-69), 2021.