

Redundancy in Manual System Testing

Blessing and Curse

2013/11/18

Benedikt Hauptmann

hauptmab@in.tum.de

Technische Universität München, Germany

Manual system tests

	Step Description	Expected Result
Step 1	Navigate: Reports -> Execute Reports -> Risk Capital Modeling	Mask 'Reporting' is shown.
Step 2	Enter: <ul style="list-style-type: none"> - Year of History = 14 - Include excluded CY = NO - Default Rating = empty Click on Button "Excel".	The years 1998-2009 are selected. A report is opened. The following tabs are generated: <ul style="list-style-type: none"> - 'Summary' - 'TE' - 'SN' - 'Fac' - 'Risk' - 'XL'
Step 3	Check the data sheet 'Summary' if the following is generated: <ul style="list-style-type: none"> - currency - time of generation - Scaling 	The correct values are shown.
Step 4	Check the data sheet 'TE' if the following columns are generated: <ul style="list-style-type: none"> - Group ID (text) - Group (text) - P.D. ID (text) 	The columns exist and are filled.

Test suites are huge

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Clone detection in manual system tests

We consider a test clone as:

- A sequence of min. 30 equal words
- No word stemming, nor stop word removal

Gaps in clones:

- up to 3 gaps
- gaps max. 10% of the overall clone length

Clone detection

- automatic clone detection based on ConQAT (www.conqat.org)
- manual inspections

```

Reports -> Execute Reports
-> Risk Capital Modeling

Mask 'Reporting' is shown.

Enter:
- Year of History = 14
- Include excluded CY = NO
- Default Rating = empty
Click on Button "Excel".

The years 1999-2012 are selected.
A report is opened. The following tabs are generated:
- 'Summary'
- 'TE'
- 'SN'
- 'Fac'
- 'Risk'
- 'XL'

Check the data sheet 'Summary' if the following is generated:
currency
time of generation
Scaling

The correct values are shown.

Check the data sheet 'TE' if the following columns are generated:
- Group ID (text)
- Group (text)
- BuPa ID (text)
- Cedent (text)
- CARAT Scenario ID (text)
- Prop Exposure PML (int)

The columns exist and are filled.
    
```

```

Reports -> Execute Reports
-> Risk Capital Modeling

Mask 'Reporting' is shown.

Enter:
- Year of History = 7
- Include excluded CY = NO
- Default Rating = empty
Click on Button "Excel".

The years 2006-2012 are selected.
A report is opened. The following tabs are generated:
- 'Summary'
- 'TE'
- 'SN'
- 'Fac'
- 'Risk'
- 'XL'

Enter:
- Year of History = empty
- Include excluded CY = NO
- Default Rating = empty
Click on Button "Excel".

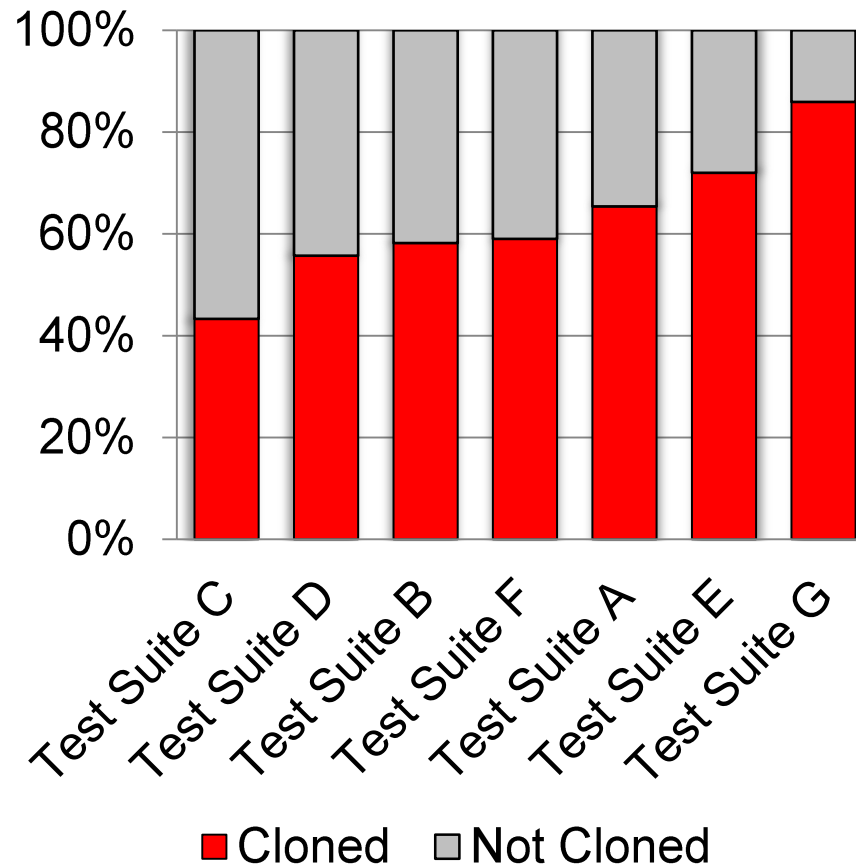
The years 1900-2012 are selected.
A report is opened. The following tabs are generated:
- 'Summary'
- 'TE'
- 'SN'
- 'Fac'
- 'Risk'
- 'XL'

Check the data sheet 'Summary' if the following is generated:
currency
time of generation
    
```

Study objects

	System under Test	Test Suite (Manual System Tests)		
		# Tests	Length (lines)	Length (words)
System A	330 kLoC	266	37,027	79,114
System B	580 kLoC	1,059	165,547	346,135
System C	150 kLoC	72	12,918	27,450
System D	430 kLoC	180	67,598	102,991
System E	760 kLoC	1,804	307,760	529,122
System F	1,400 kLoC	135	22,903	34,136
System G	160 kLoC	605	127,385	317,205

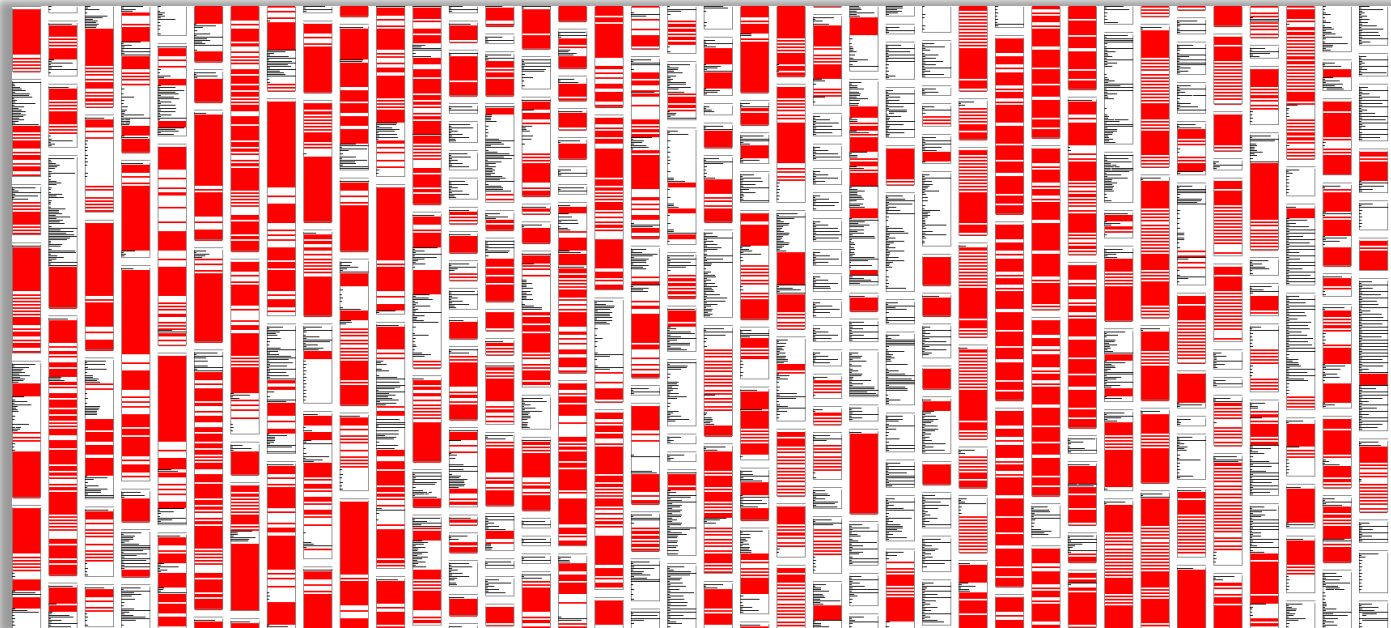
Amount of cloning in the study objects



How can we use this information?

1. Remove all clones

nice idea, but will cause tremendous effort (very unrealistic)



How can we use this information?

1. Remove all clones

nice idea, but will cause tremendous effort (very unlikely)

2. Avoid new clones

requires continuous clone control



How can we use this information?

1. Remove all clones

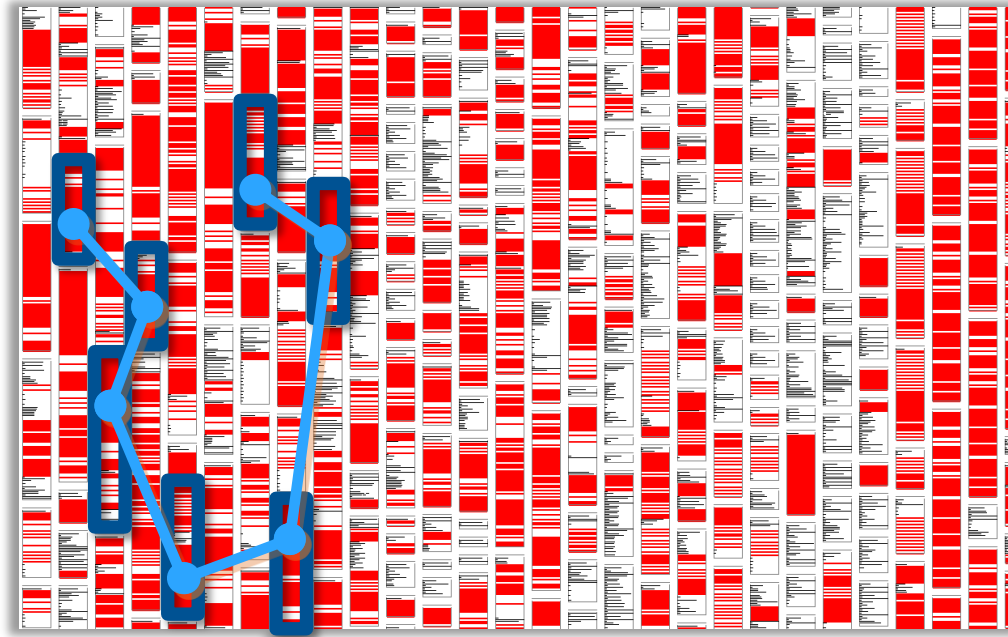
nice idea, but will cause tremendous effort (very unlikely)

2. Avoid new clones

requires continuous clone control

3. Guide test engineers during maintenance tasks

apply patches consistently



How can we use this information?

1. Remove all clones

nice idea, but will cause tremendous effort (very unlikely)

2. Avoid new clones

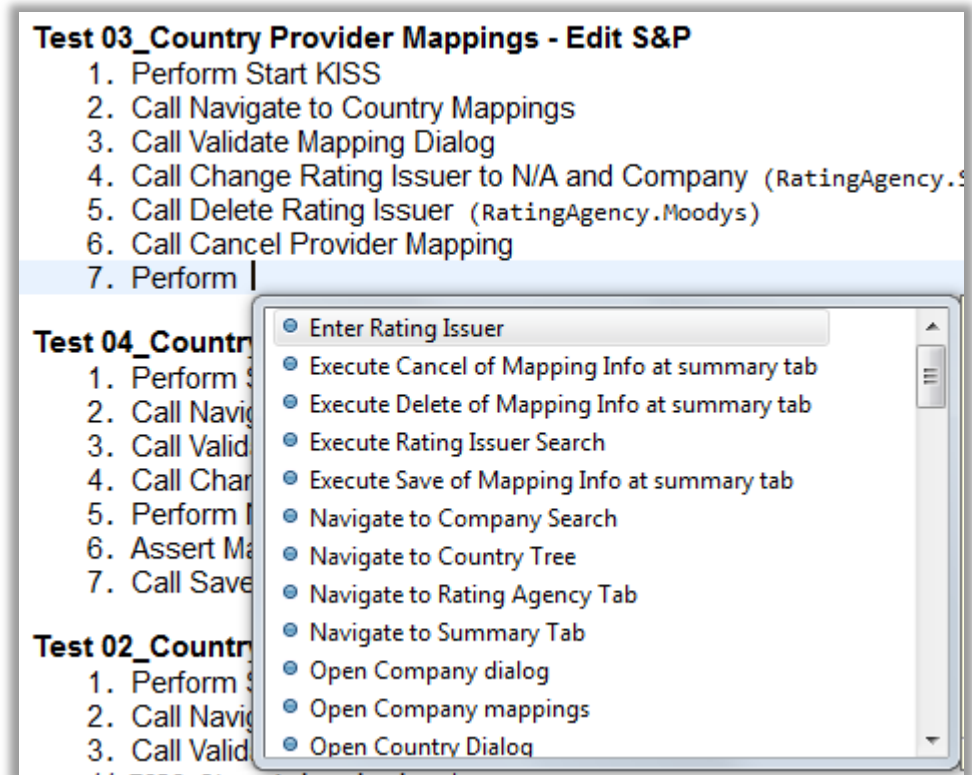
requires continuous clone control

3. Guide test engineers during maintenance tasks

apply patches consistently

4. Input for testing tool selection

improve reuse techniques
of testing languages/tools



The screenshot displays a list of test cases and a dropdown menu. The test cases are:

- Test 03_Country Provider Mappings - Edit S&P**
 1. Perform Start KISS
 2. Call Navigate to Country Mappings
 3. Call Validate Mapping Dialog
 4. Call Change Rating Issuer to N/A and Company (RatingAgency.S
 5. Call Delete Rating Issuer (RatingAgency.Moodys)
 6. Call Cancel Provider Mapping
 7. Perform I
- Test 04_Country**
 1. Perform S
 2. Call Navig
 3. Call Valid
 4. Call Char
 5. Perform I
 6. Assert Ma
 7. Call Save
- Test 02_Country**
 1. Perform S
 2. Call Navig
 3. Call Valid

The dropdown menu is open and shows the following options:

- Enter Rating Issuer
- Execute Cancel of Mapping Info at summary tab
- Execute Delete of Mapping Info at summary tab
- Execute Rating Issuer Search
- Execute Save of Mapping Info at summary tab
- Navigate to Company Search
- Navigate to Country Tree
- Navigate to Rating Agency Tab
- Navigate to Summary Tab
- Open Company dialog
- Open Company mappings
- Open Country Dialog

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

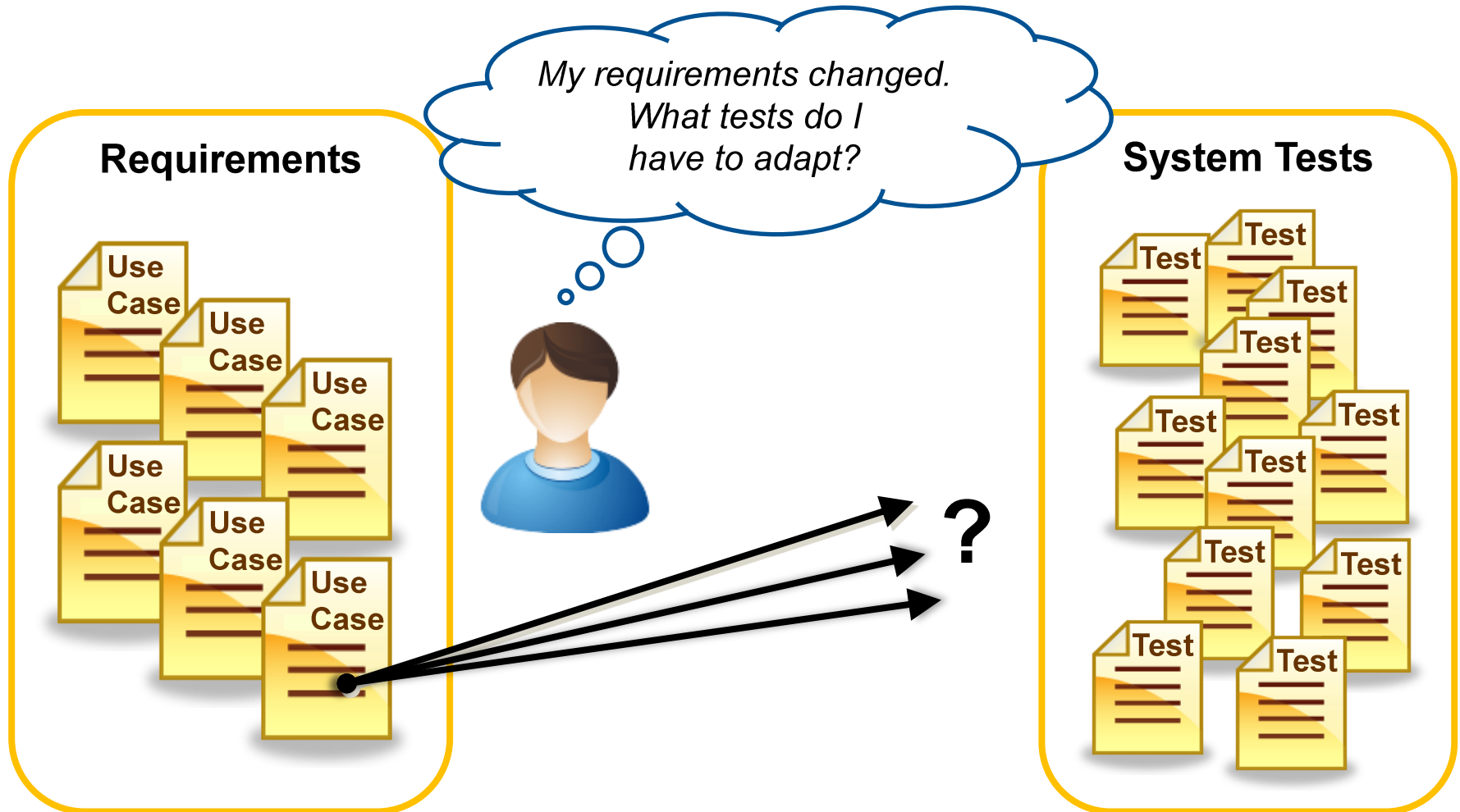
- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

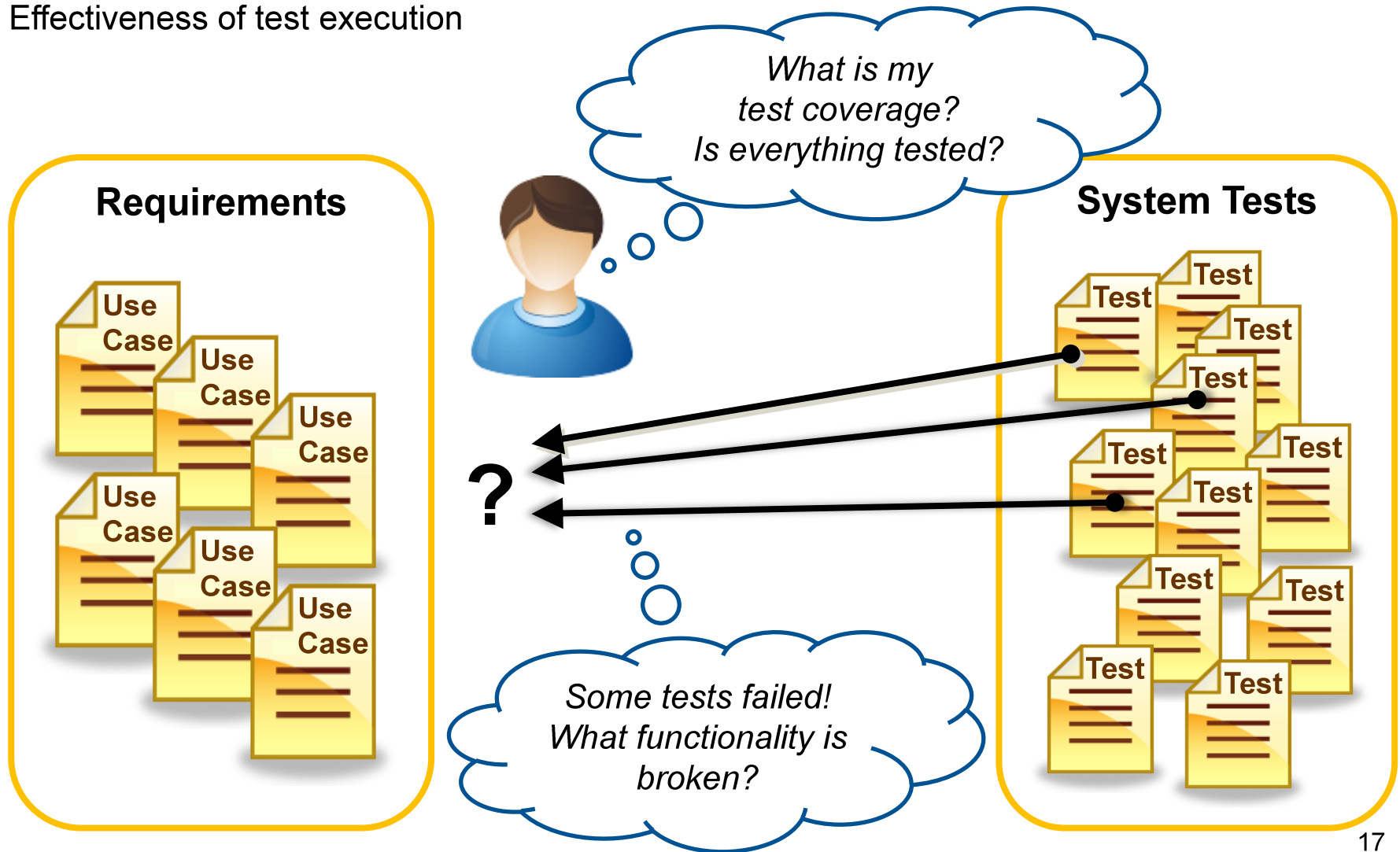
Tracing between tests and requirements

Maintenance of test suites



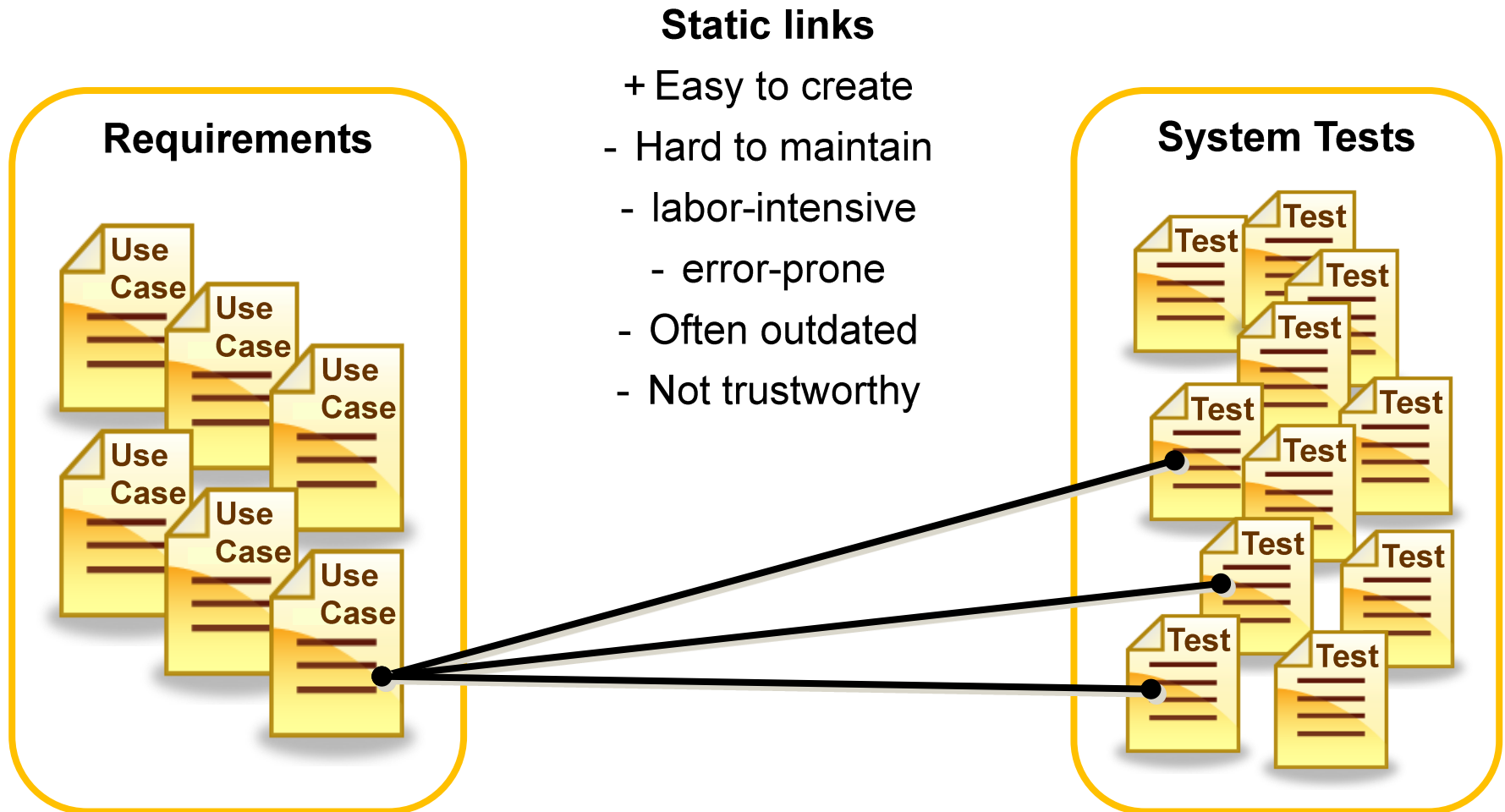
Tracing between tests and requirements

Effectiveness of test execution



Tracing between tests and requirements

A common approach



Tracing between tests and requirements

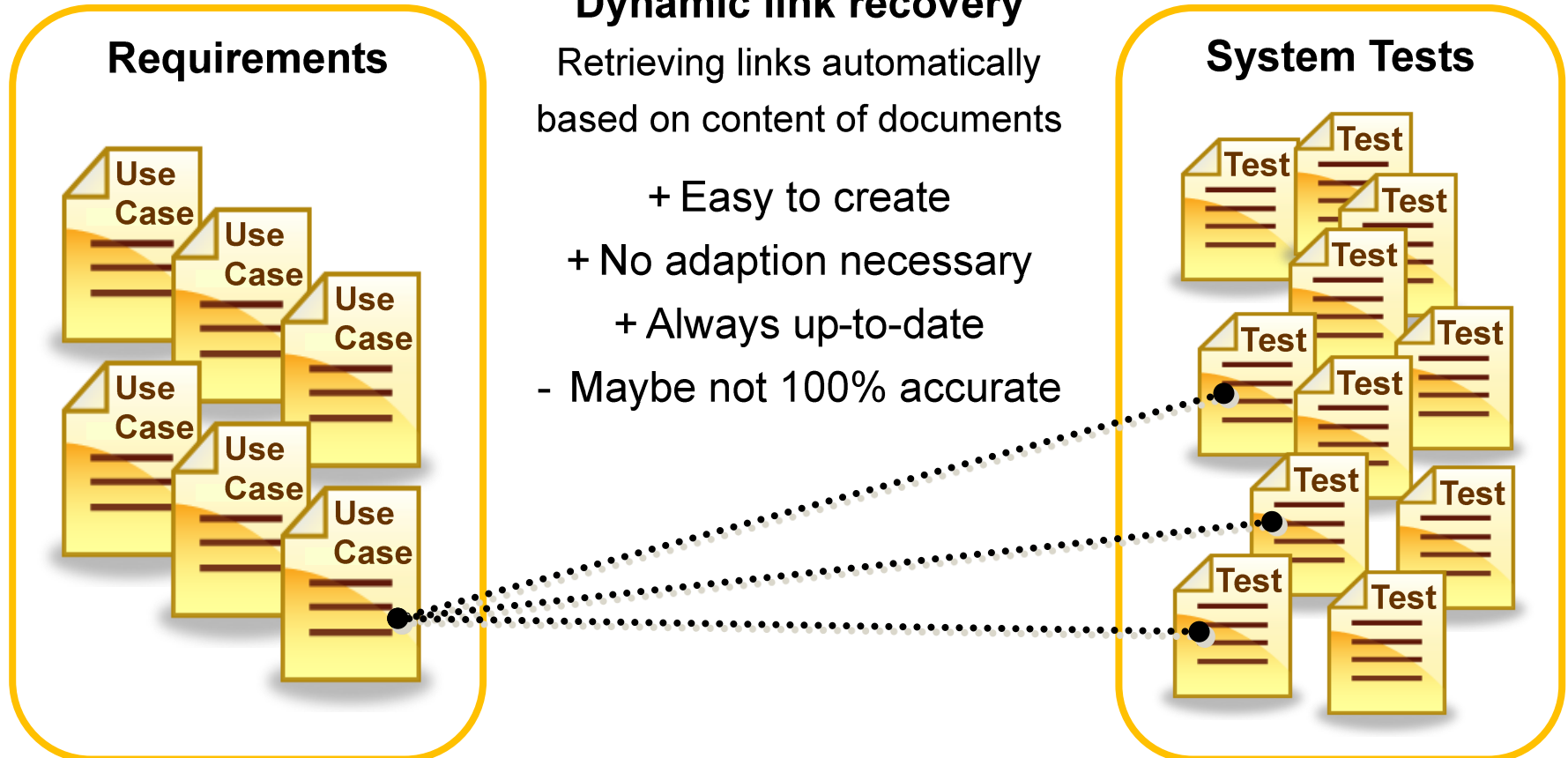
More promising ...

Approach:

Dynamic link recovery

Retrieving links automatically based on content of documents

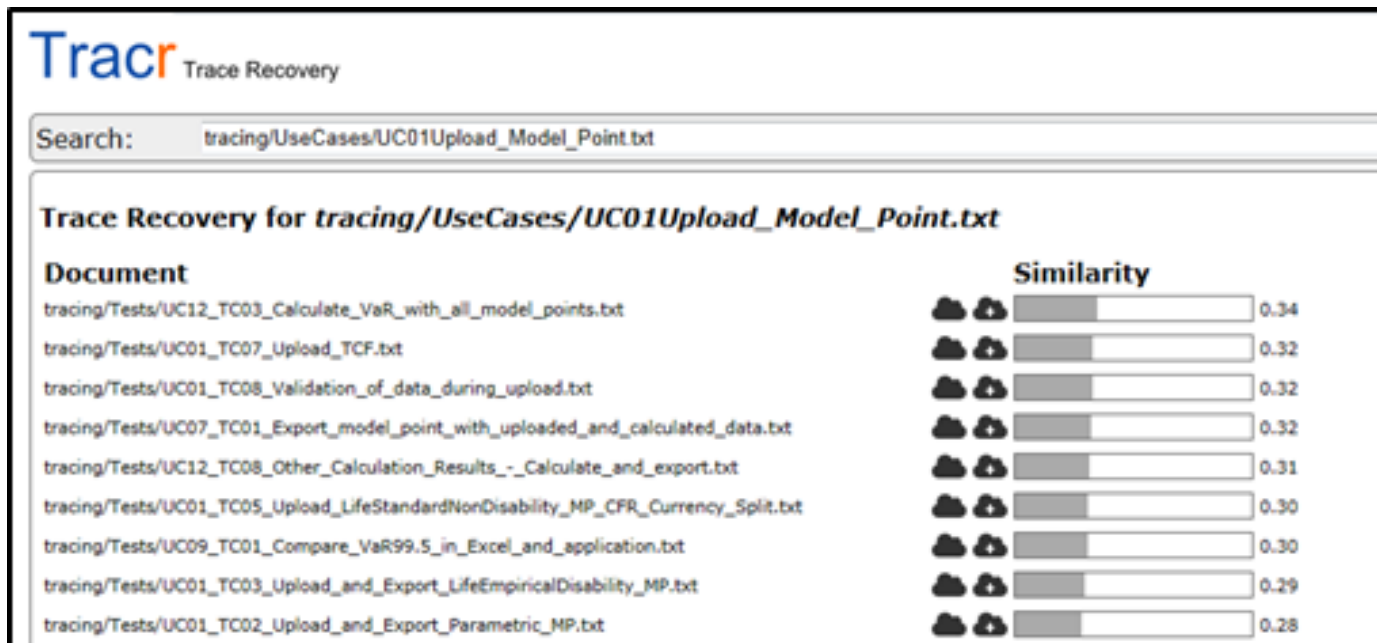
- + Easy to create
- + No adaption necessary
- + Always up-to-date
- Maybe not 100% accurate



Tracing between tests and requirements

Dynamic link recovery - implementation

- Calculation a *similarity value* for each pair of use case and test case using *Latent Semantic Indexing (LSI)*
(Identifies patterns in relationships between terms in unstructured text)
- Allows tracing in both directions



Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Some examples of quality defects

repeatability?

Step ...	Step Description	Expected Result
Step 6	Repeat step 4 (Tab Section/Search) and step 5 as long as you want.	...

comprehensibility?

Step	Step Description	Expected Result
Step 12
Step	The entries differ depending on chosen View Mode but ...

maintenance?

The collage shows multiple overlapping screenshots of a software manual. Each screenshot contains a table with columns for 'Step', 'Step Description', and 'Expected Result'. The text in the screenshots is mostly illegible due to the overlapping and low resolution, but some fragments are visible, such as 'Reports -> Execute Report -> Risk Capital Modeling' and 'Years of History = 14'. The red annotations 'repeatability?', 'comprehensibility?', and 'maintenance?' are overlaid on these screenshots to highlight specific quality defects.

Test smells for tests in natural language

Automatically Measurable Natural Language Test Smells

1. Hard-Coded Values
2. Long Test Steps
3. Conditional Tests
4. Badly Structured Test Suite
5. Test Clones
6. Ambiguous Tests
7. Inconsistent Wording

Inspired by

- Code smells
- (Unit) test smells
- Own experiences from industry

**Every smell is
automatically
measurable**

Test smell *Inconsistent Wording* – example

Smell *Inconsistent Wording*:

Domain concepts are not used in a consistent way (e.g., several names are used for the same domain concept).

→ *Comprehension*: It is difficult to detect similarities of tests.

	Step Description	Expected Result		
...	Expected Result
Step 2	The following countries are selectable: Germany, USA, Spain
Step 3	Select Germany as licensed country for the selected contract.	...	Step 16	Make sure that all listed contracts are assigned to a valid area.
...
...

Test smell *Inconsistent Wording* – example

Smell *Inconsistent Wording*:

Domain concepts are not used in a consistent way (e.g., several names are used for the same domain concept).

→ *Comprehension*: It is difficult to detect similarities of tests.

Automatic detection:

1. Clean text

	Step Description	Expected Result		
...	Expected Result
Step 2	The following countries are selectable: Germany, USA, Spain
Step 3	Select Germany as licensed country for the selected contract.	...	Step 16	Make sure that all listed contracts are assigned to a valid area.
...

Test smell *Inconsistent Wording* – example

Smell *Inconsistent Wording*:

Domain concepts are not used in a consistent way (e.g., several names are used for the same domain concept).

→ *Comprehension*: It is difficult to detect similarities of tests.

Automatic detection:

1. Clean text
2. Find synonym groups
 - Thesauri
 - Project glossary

	Step Description	Expected Result
...
Step 2	The following country s are selectable: Germany, USA, Spain	...
Step 3	Select Germany as licensed country for the selected contract.	...
...

	Expected Result
...	...
Step 16	Make sure that all listed contracts are assigned to a valid area .
...	...

Test smell *Inconsistent Wording* – example

Smell *Inconsistent Wording*:

Domain concepts are not used in a consistent way (e.g., several names are used for the same domain concept).

→ *Comprehension*: It is difficult to detect similarities of tests.

Automatic detection:

1. Clean text
2. Find synonym groups
 - Thesauri
 - Project glossary
3. Selecting the ‘right’ word
 - Find most common synonym
 - Project glossary

	Step Description	Expected Result
...
Step 2	The following country s are selectable: Germany, USA, Spain	...
Step 3	Select Germany as licensed country for the selected contract.	...
...

	Expected Result
...	...
Step 16	Make sure that all list contracts are assigned to a valid area .
...	...

Agenda

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent usage of synonyms
- Level of redundancy: semantic

Redundancy in Manual System Testing – Blessing and Curse

Overall goal: measure and improve the quality of natural language tests

Focus on: creating, executing, maintaining and understanding tests

1. Test clones

- Utilization of redundancy: find similar parts of tests
- Level of redundancy: syntax

2. Tracing requirements and tests

- Utilization of redundancy: recover links between requirements and tests
- Level of redundancy: semantic

3. Test smell: inconsistent synonyms

- Utilization of redundancy: find inconsistent eg. on synonyms
- Level of redundancy: semantic

Thank you :-)

CURSE
BLESSING
CURSE




Benedikt Hauptmann
hauptmab@in.tum.de
www4.in.tum.de/~hauptmab
Technische Universität München