Automated application's user interface visual testing using static analysis approach

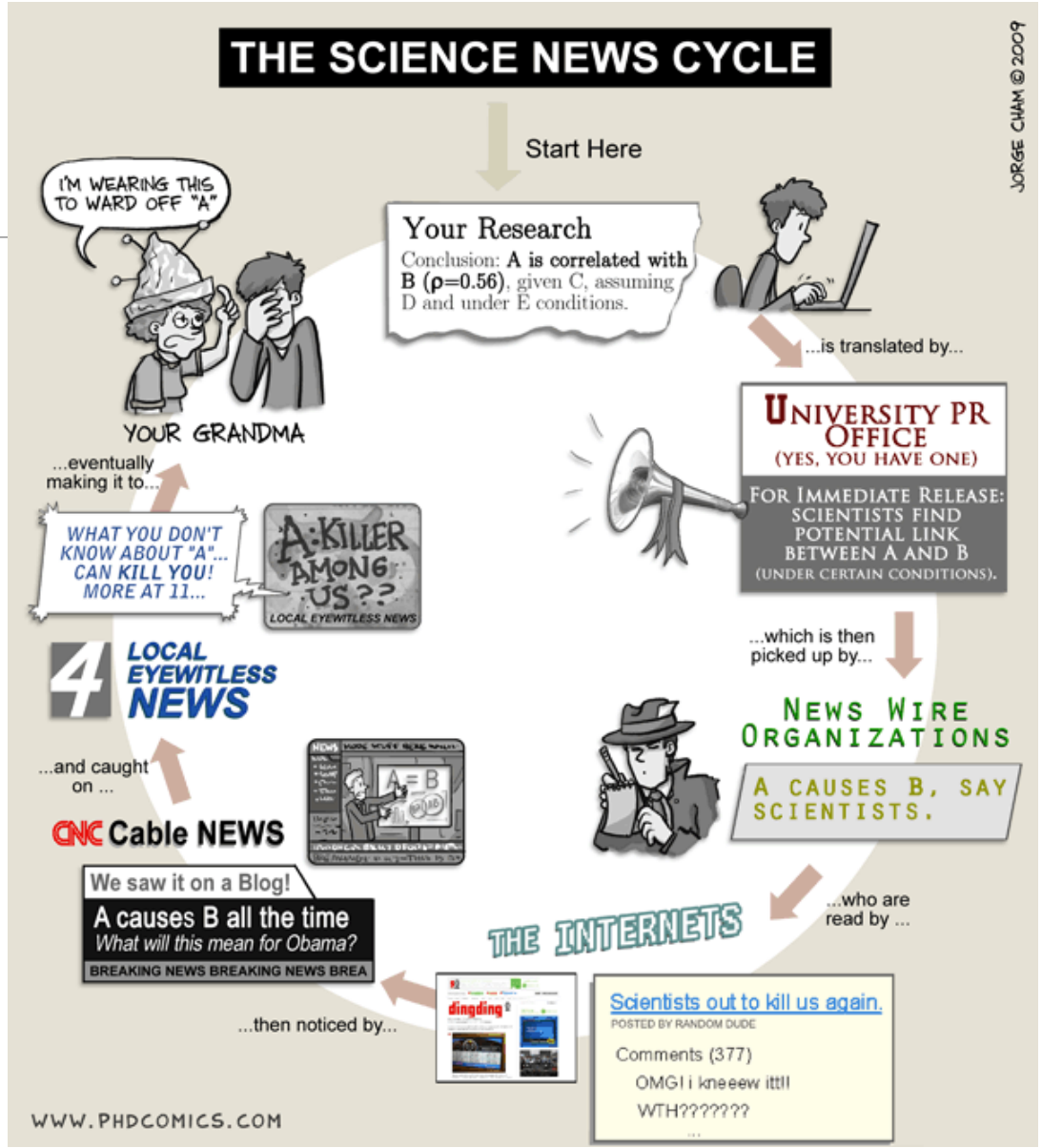
ŠARŪNAS PACKEVIČIUS ET AL. @ KTU

Research

Small increments

Specialized

AI is here to take our jobs



Research topics/what we do

Overcoming the Equivalent Mutant Problem: A Systematic Literature Review and a Comparative Experiment of Second Order Mutation

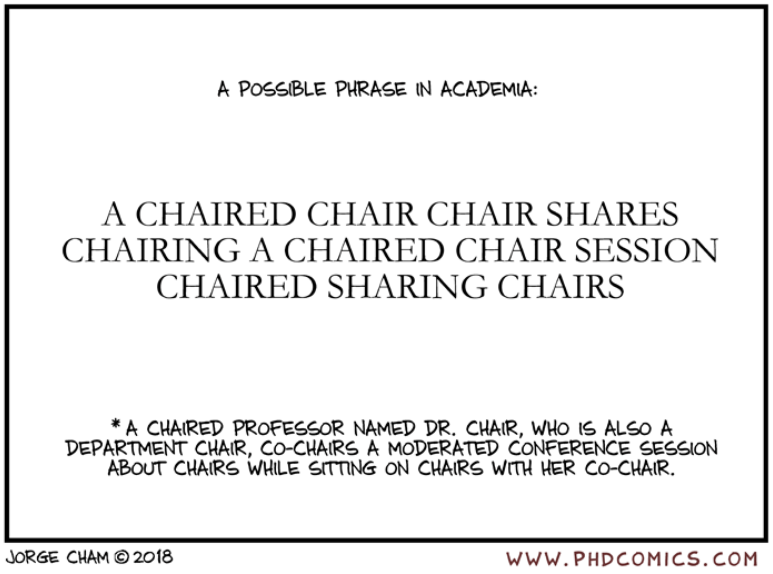
GenProg: A Generic Method for Automatic Software Repair

A Systematic Literature Review of Software Defect Prediction: Research Trends, Datase Methods and Frameworks

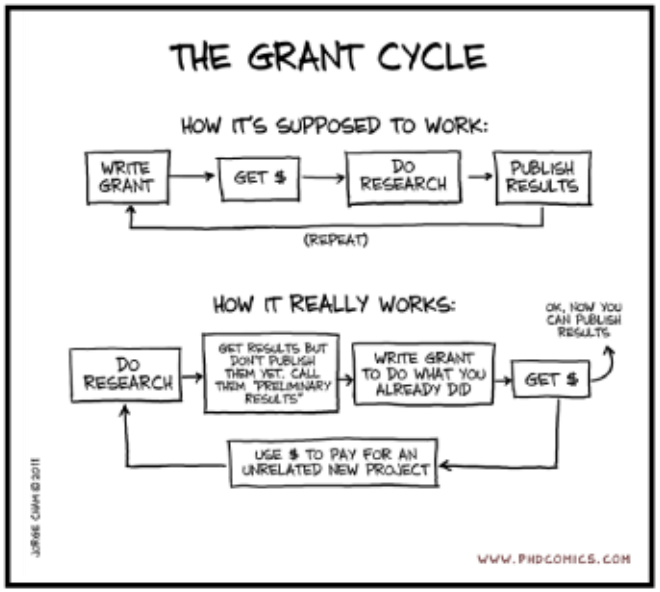
Using Natural Language Processing to Automatically Detect Self-Admitted Techni Debt

Automated Checking of Conformance to Requirements Templates Using Natural Lan Processing

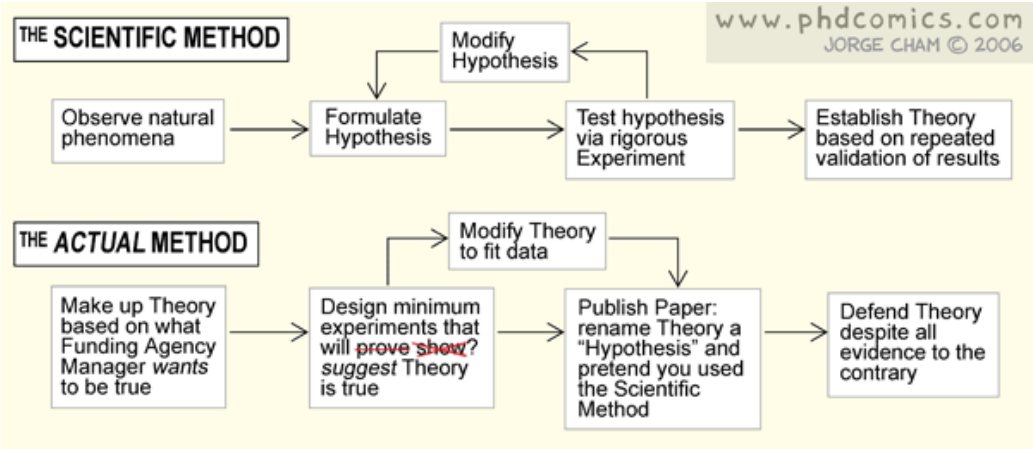
Text Filtering and Ranking for Security Bug Report Prediction



And how we do it



"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com



When I can get it?

Test-driven development

- 12000 BC : OgTheVenerable, holding a raw piece of meat, describes the process of creating fire to his fellow cavemen, setting their expectations and a prescription for his own task. The piece of charred meat he held in his hand afterward is widely regarded as the first passing test case.
- 1959 through 1963 : programmers for the Mercury Space Program use a form of TDD while programming punched cards

Scrum

- Hirotaka TakeuchiIkujiro Nonaka, The New New Product Development Game, *Harvard Business Review*, January 1986

Well…

Research trends

Automatic Software Repair

The Oracle Problem in Software Testing

Software Fault Localization

Defect Prediction

Technical Debt

Persistent Software Errors

Software Defect Prediction: Research Trends, Datasets, Methoc

1. **Estimation**
2. **Association**
3. **Classification**
4. **Clustering**
5. **Dataset Analysis**

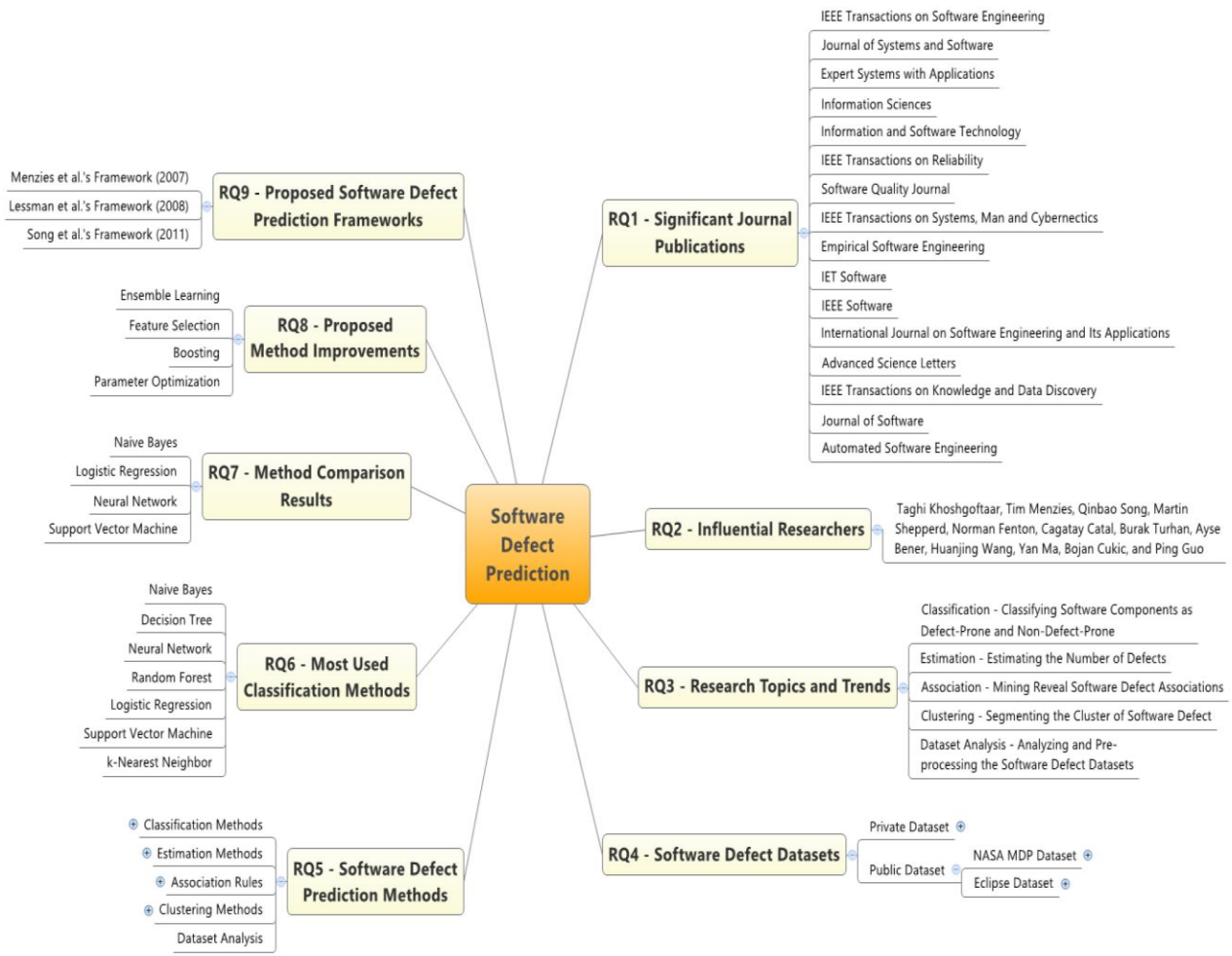


Figure 16 Complete Mind Map of the SLR on Software Defect Prediction

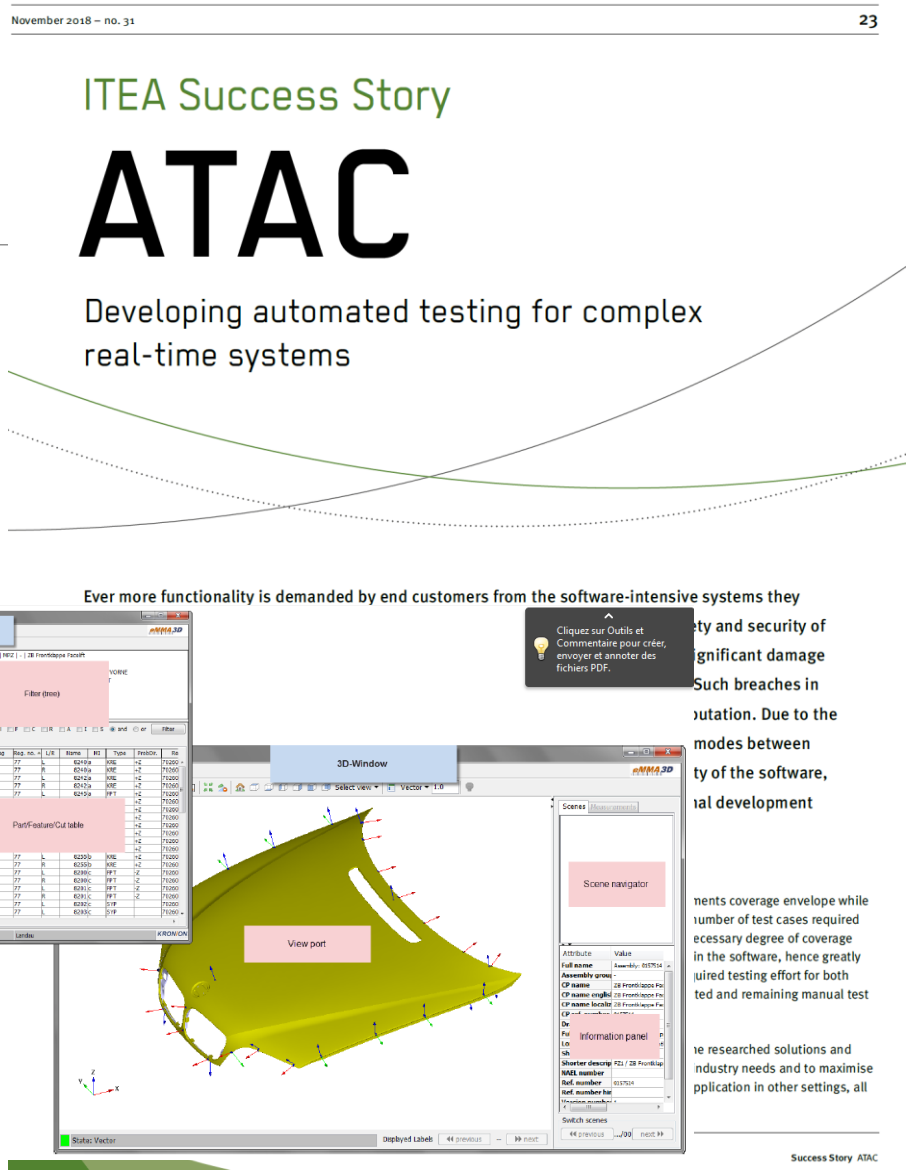
KTU Topics

Model Based Testing

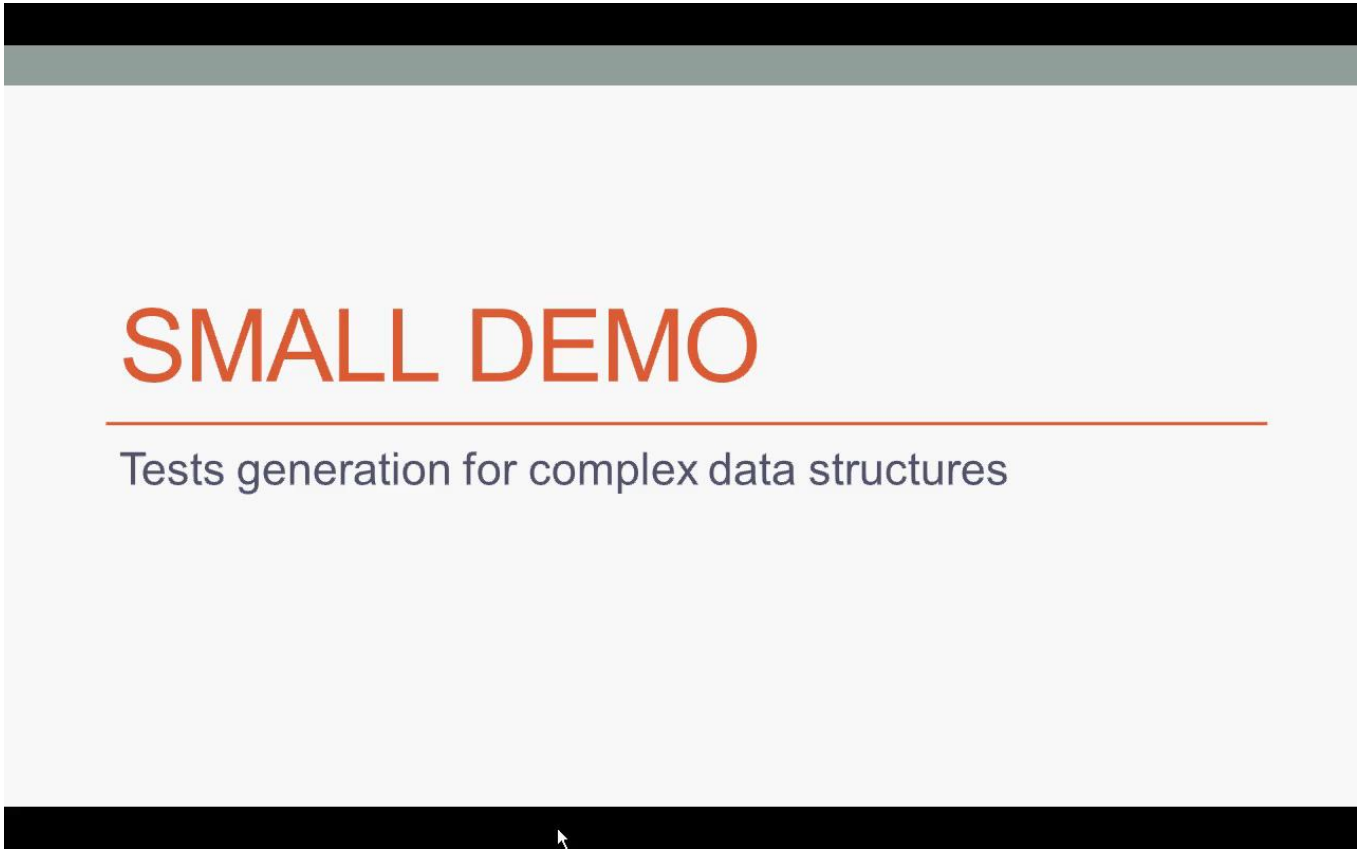
GUI Defect Search

Hardware testing

- FPGA, VHDL, ….



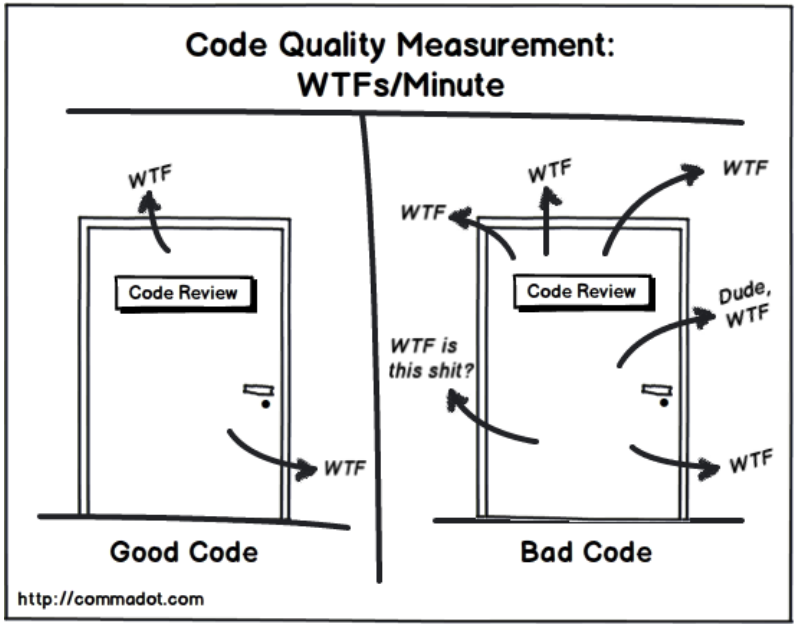
Model Based Testing



GUI Testing

Have you ever looked at the app at thought: something does not look right here….

Were you able to objectively describe what's wrong and how should it be?



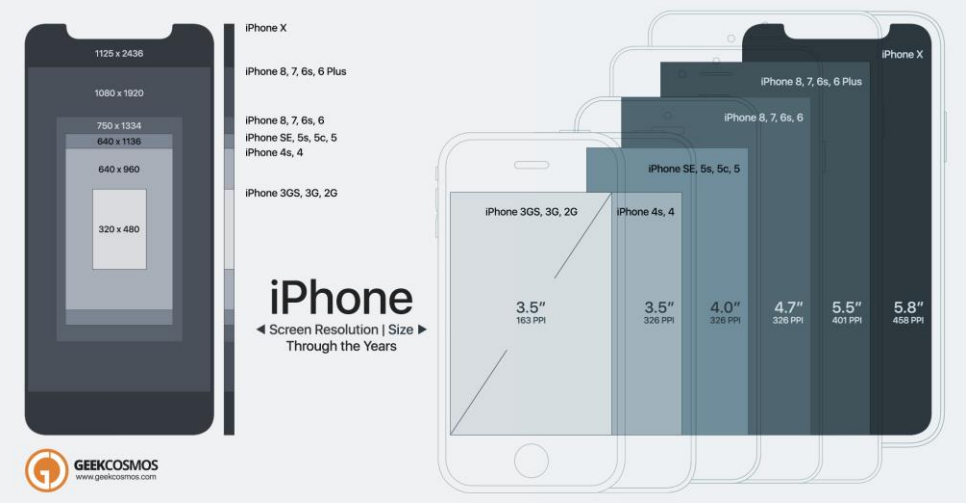
Mobile app “features” - problems

Various device configurations – “well it works on mine, should be ….”

- Bad layout, missing, invisible text.
- Unreadable texts

Internationalization

- Spelling, grammar errors
- Automated translation
- Poor text style
- Messages are not always “understandable”



User interface defects

Layout problems:

- Invisible control
- Clipped control
- Too small control
- Misaligned controls

Text problems:

- Too small text
- Partial text

Image problems:

- Bad scaling
- Bad resolution

Colour problems:

- Hidden control
- Not matching colours
- Poor colour choice

Application menu problems:

- Menu does not fit to a screen/toolbar
- Missing icons.

Navigation problems:

- Stuck at screen
- Wrong window
- App crash
- Timing problems

Experiment

Test 1000 apps.

Log defects

Clasify defects

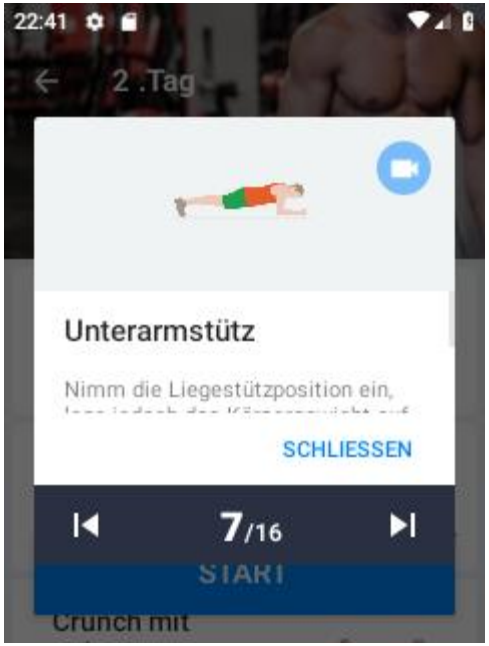
Define automated testing method

Define defect detection rules

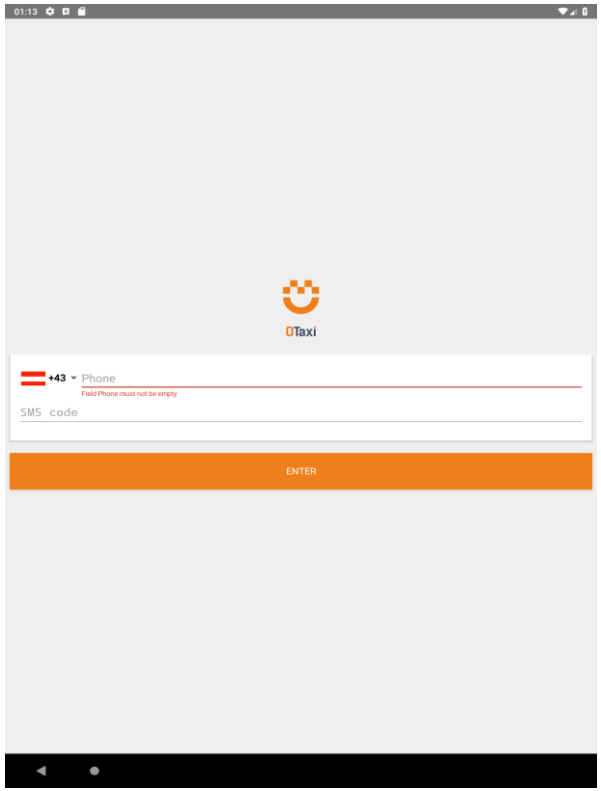
Run rules on 1000 apps.

Compare results – prove that it's good.

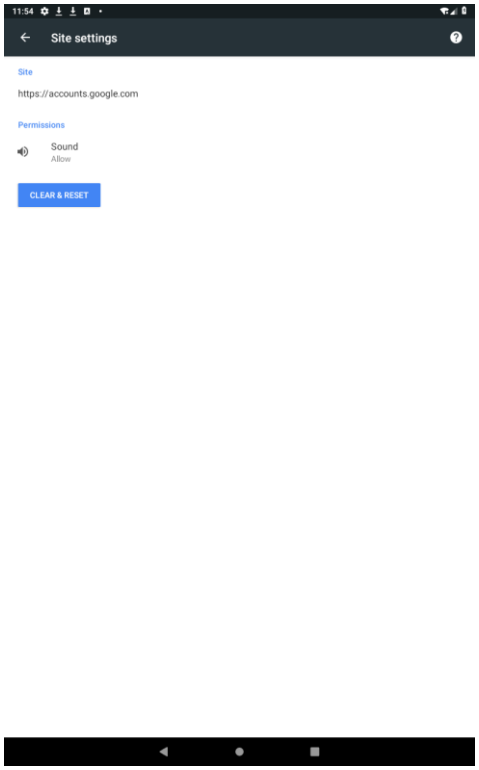
Defects


22:41
2 .Tag
Unterarmstütz
Nimm die Liegestützposition ein,
SCHLIESSEN
7/16
STARI
Crunch mit


14:05
TELE2
+370 XXXXXXXX
Užpildykite tuščią telefono numerio lauką
Prisiminti prisijungimo
REGISTRACIJA
PRISIJUNGTI


01:13
OTaxi
+43
Phone
Field Phone must not be empty
SMS code
ENTER

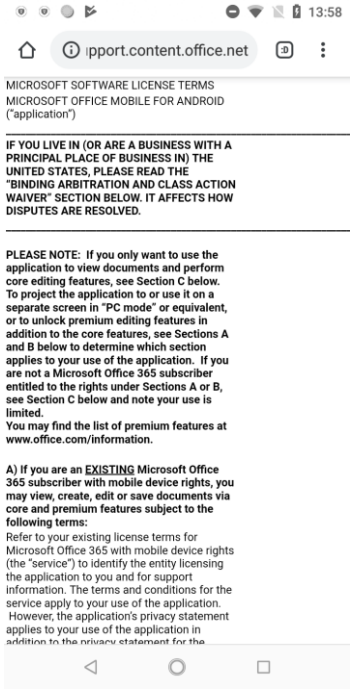
Defects…

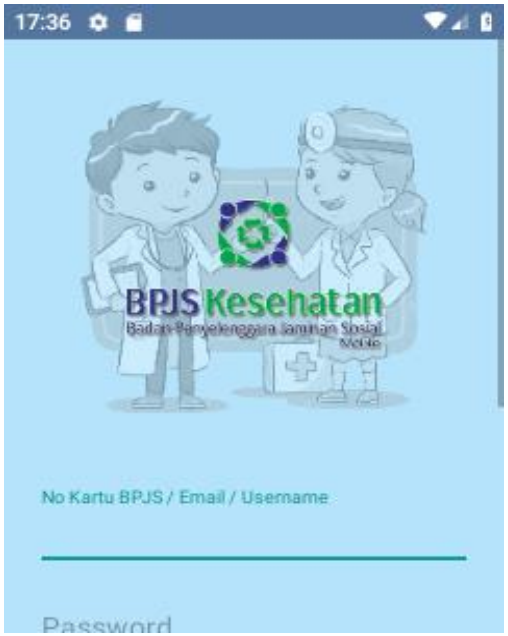

11:54
Site settings
Site
https://accounts.google.com
Permissions
Sound
Allow
CLEAR & RESET

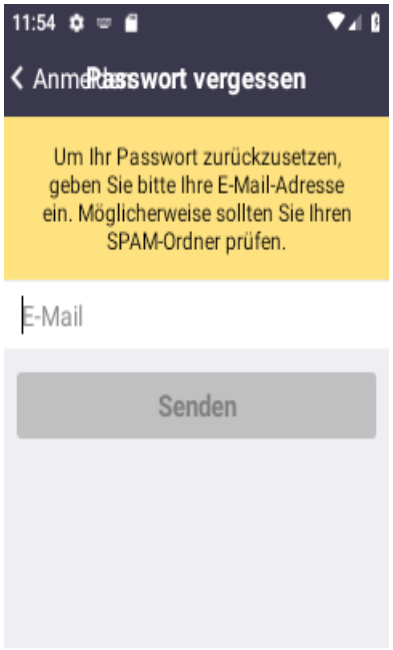

21:12
Kilométrage
Distance
Kilomètres
Essence
Gallons
Gallons au 100 milles


22:01
Currency Converter
CONVERTER
CURRENCIES
NEWS
TWITTER
EUR - Euro
USD - United States Do..
1
1,40
Rate : 1.4000 (0 Seconds ago)
Test Ad

Defects


13:58
MICROSOFT SOFTWARE LICENSE TERMS
MICROSOFT OFFICE MOBILE FOR ANDROID
("application")
IF YOU LIVE IN (OR ARE A BUSINESS WITH A PRINCIPAL PLACE OF BUSINESS IN) THE UNITED STATES, PLEASE READ THE "BINDING ARBITRATION AND CLASS ACTION WAIVER" SECTION BELOW. IT AFFECTS HOW DISPUTES ARE RESOLVED.
PLEASE NOTE: If you only want to use the application to view documents and perform core editing features, see Section C below. To project the application to or use it on a separate screen in "PC mode" or equivalent, or to unlock premium editing features in addition to the core features, see Sections A and B below to determine which section applies to your use of the application. If you are not a Microsoft Office 365 subscriber entitled to the rights under Sections A or B, see Section C below and note your use is limited.
You may find the list of premium features at www.office.com/information.
A) If you are an EXISTING Microsoft Office 365 subscriber with mobile device rights, you may view, create, edit or save documents via core and premium features subject to the following terms:
Refer to your existing license terms for Microsoft Office 365 with mobile device rights (the "service") to identify the entity licensing the application to you and for support information. The terms and conditions for the service apply to your use of the application. However, the application's privacy statement applies to your use of the application in


17:36
BPJS Kesehatan
No Kartu BPJS / Email / Username
Password


11:54
Passwort vergessen
Um Ihr Passwort zurückzusetzen, geben Sie bitte Ihre E-Mail-Adresse ein. Möglicherweise sollten Sie Ihren SPAM-Ordner prüfen.
E-Mail
Senden

The proposed method

1. Execute app on many devices.

2 .Take screenshots at each execution step.

3. Search for defects in each screenshot:
   - Defined defect classes.
   - Defect detection rules per each class.

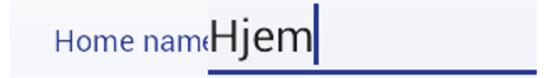
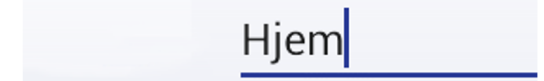
Detected Defects Summary

Defect Type	English	Non-English	Phone	Tablet	Small Screen	Normal Screen	Large Screen	Total
Bad Spelling	1	10	8	3	7	1	3	11
Wasted Space	208	2294	711	1791	546	83	1873	2502
Technical Jargon	260	488	589	159	425	98	225	748
Clipped Text	1842	4426	5800	468	4302	1035	931	6268
Wrong Font Sizes	0	91	3	88	3	0	88	91
Untranslated Text	176	1075	1059	192	550	469	232	1251
Bad Colors	73	210	192	91	118	50	115	283
Bad Scaling	37	284	150	171	128	12	181	321
Low Res Image	8	63	53	18	26	24	21	71
Clipped Control	359	1165	1303	221	1105	94	325	1524
No Margins	120	169	256	33	181	65	43	289
Uncentered	104	323	323	104	278	35	114	427
Unfilled Placeholder	0	30	21	9	17	4	9	30
Bad Margins	194	277	394	77	277	45	149	471
Obscured Control	36	159	193	2	183	9	3	195
No Anti Aliasing	0	6	6	0	6	0	0	6
Empty View	759	2274	2307	726	1115	899	1019	3033
Unreadable Text	19	305	310	14	280	25	19	324
Clashing Background	36	101	110	27	64	37	36	137
Unaligned Controls	92	162	247	7	223	21	10	254
Crowded Controlls	18	89	105	2	105	0	2	107
Obscured Text	115	413	509	19	428	38	62	528
Unlabeled Entry Field	171	349	399	121	234	86	200	520
Not Enough Space	34	285	257	62	205	45	69	319
Missing Text	29	33	43	19	29	4	29	62
Wrong Encoding	0	3	2	1	1	1	1	3
Misaligned Control	0	25	1	24	0	1	24	25
Too Large Control	0	2409	0	2409	0	0	2409	2409

Hjem

Classification of the user-interface text defects

Text presentation defects

No	Type	Example
TP1	Text placement	
TS1	Font sizes	
TS2	Unreadable text	
TB1	Clashing background	
TC1	Partial text	
TC2	Clipped text	
TE1	Wrong encoding	
TM1	Missing text	

Classification of the user-interface text defects

Text semantics defects

No	Type	Example
SD1	Synonyms	*The schedule* is assigned. *The week program* is not set.
SD2	Wrong terminology	*Loyalty card* is expired. (instead of a "*gift card*")
SD3	Unclear terminology	Heater is off. (meaning "electric one") Heater is on. (meaning "thermostat")
SD4	Jargon	Establishing connection to remote web socket.
SS1	Bad text style	press und hold button for fvie sek chek red LED.
SS2	Too much text	No program is assigned to this zone. To program this zone, you should either create a new program or select an existing program from the list below.
SS3	Vague error messages	Connection was reset by peer
SS4	Untactful messages	Error!!! Invalid input fields. Enter correct data!!!
SS5	Misleading text	Device firmware does not support this method, please use the first method
SU2	Too difficult to under-stand the text	As many automated test input generation tools for Android need to instrument the system or the app, they cannot be used in some scenarios such as compatibility testing and malware analysis.
SL1	Missing translations	error_message_no_wifi
SL2	Wrong language	You could reset device to factory defaults and try again. Trykk og hold nede [+], [-] and [OK] knapper på gulvvarmetermostaten i 5 sekunder.

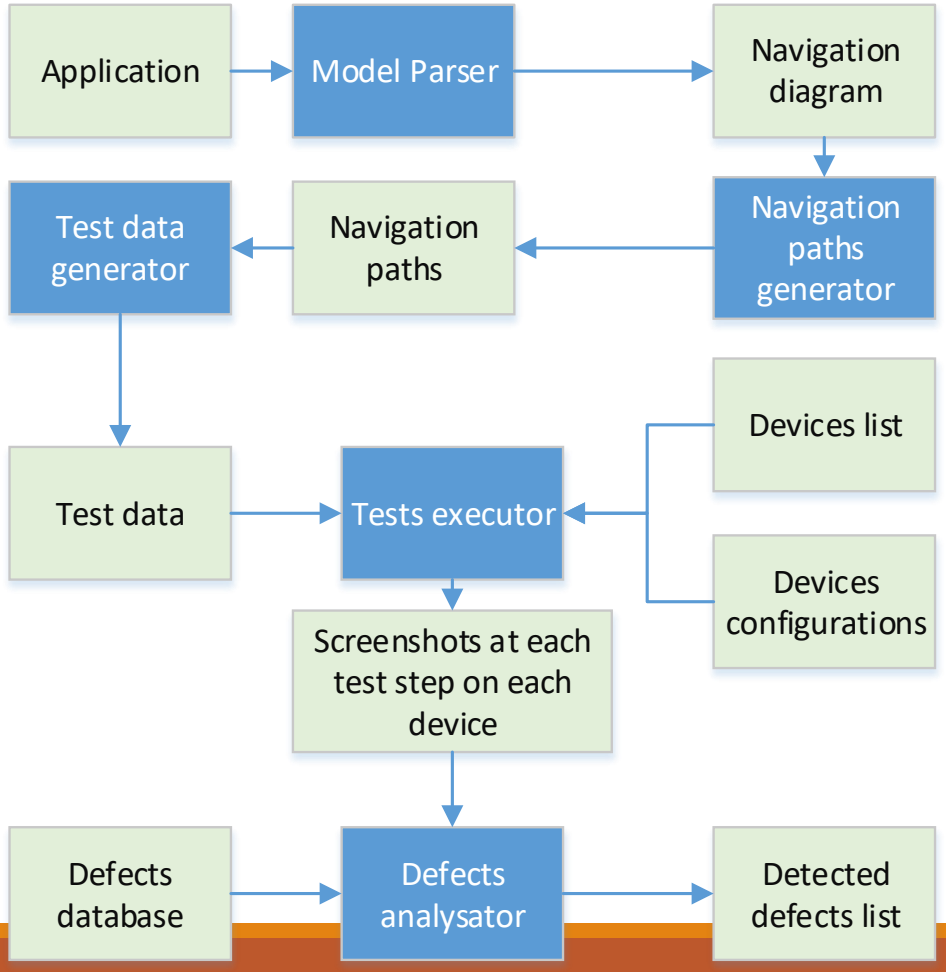
Detection of the defects

Sources:

- Screenshots
- Application window texts
- Application texts

Process:

- Generate app execution paths.
- Run on many devices.
- Take screenshots.
- Analyse all screenshots



Detection rules

Defect Type	Algorithms
	Input: *recognizedTexts*, recognized texts **Input:** *textAreas*, recognized texts areas **Input:** *actualTexts*, actual window texts in current language **Input:** *i18nTexts*, actual window texts in all app languages **Input:** *window*, analysed window **Input:** *device*, device screenshot was taken **Input:** path, execution path through app **Input:** *textPlaceHolders*, placeholders for translations **Output:** *defect*, detected defect

Unreadable text	**foreach** *text* ∈ *actualTexts* **do**
	  if *recognizedTexts* **not** *contains text* **then**
	  defect ← *window*, *device*, *text*
	foreach *area* ∈ *textAreas* **do**
	  fontHeight ← *calculatePhysicalSize(textArea, device)*
	  if *fontHeight* < *2mm* **then**
	  defect ← *window*, *device*, *area*

Clashing background	**foreach** *text* ∈ *actualTexts* **do**
	  if *recognizedTexts* **not** *contains text* **then**
	  defect ← *window*, *device*, *area*
	  if |*text.color* – *window.backgroundColor*| < 5 **then**
	  defect ← *window*, *device*, *area*

Partial text	**foreach** *text* ∈ *actualTexts* **do**
	  if *recognizedTexts* **not** *contains text* **then**
	  foreach *recognizedText* ∈ *recognizedTexts* **do**
	  if *text contains substring recognizedText* **then**
	  defect ← *window*, *device*, *area*

Defect Type	Algorithms

Missing text	**foreach** *text* ∈ *actualTexts* **do**
	  if *recognizedTexts* **not** *contains text* **then**
	  defect ← *window*, *device*, *area*

Jargon	**foreach** *text* ∈ *actualTexts* **do**
	  foreach *word* ∈ *text* **do**
	  if not *isStopWord(word)* **and** *isNoun(word)*
	  and *isTechnical(word, actualTexts)* **then**
	  defect ← *window*, *device*, *text*

Bad text style	**foreach** *recognizedText* ∈ *recognizedTexts* **do**
	  if not spellCheck(*recognizedText)* **or**
	  not grammarCheck(*recognizedText)* **then**
	  defect ← *window*, *device*, *recognizedText*

Untactful messages	**foreach** *recognizedText* ∈ *recognizedTexts* **do**
	  if *isOffensive*(*recognizedText)* **then**
	  defect ← *window*, *device*, *recognizedText*

Too hard to understand text	**foreach** *text* ∈ *actualTexts* **do**
	  if *readabilityIndex(text)* ≥ *CollegeLevel* **then**
	  defect ← *text*, *window*

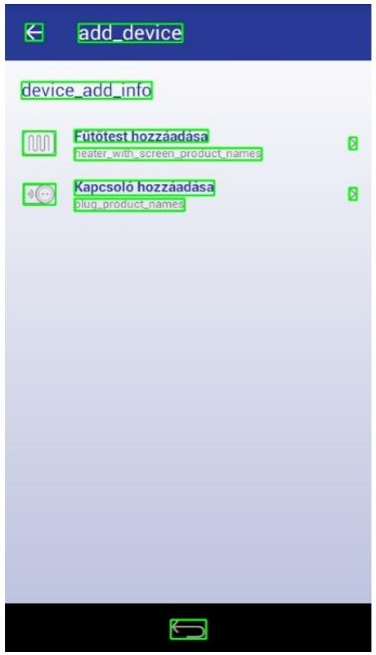
Missing translations	**foreach** *recognizedText* ∈ *recognizedTexts* **do**
	  if *textPlaceHolders contains recognizedText* **then**
	  defect ← *window*, *device*, *recognizedText*

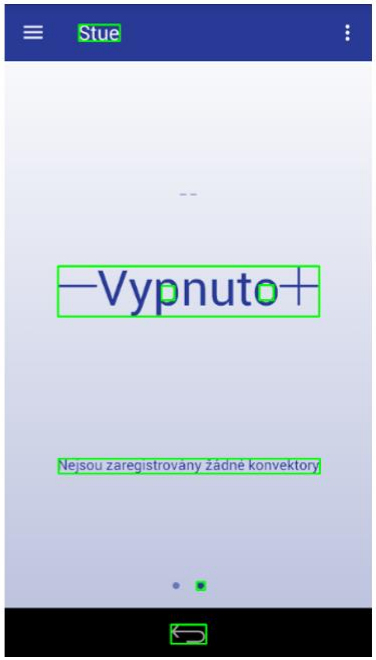
Wrong language	**foreach** *recognizedText* ∈ *recognizedTexts* **do**
	  if *actualTexts* **not** *contains recognizedText* **then**
	  if *actualTexts contains i18nTexts* **then**
	  defect ← *window*, *device*, *recognizedText*

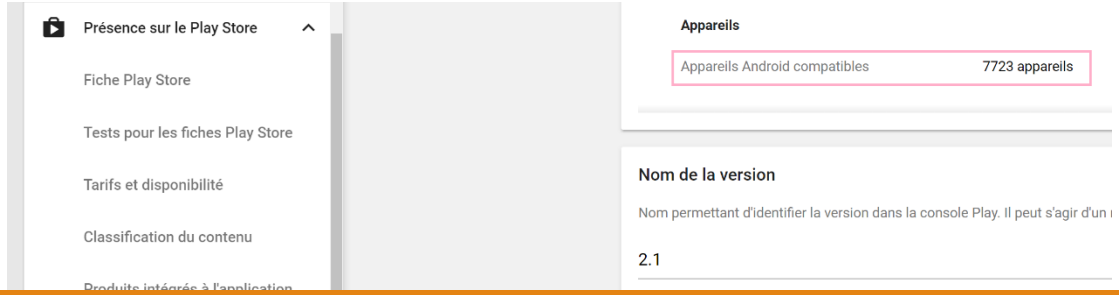
Testing example

One small app as showcase

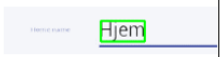
- Multilanguage support
- iPhone and Android apps
- Multiple devices support







Testing results

Defect	Device	Step/Window	Screenshot	Detected text	Expected Text
TS1	Jolla (EN)	Settings		Home name Hjem	Home name Hjem
TM1	Jolla (EN)	Settings		Hjem	Home name Hjem
TB1	Jolla (EN)	Settings		Home name Hjem	Home name Hjem
TC1	Xperia Ray	Settings		Hom name Hjem	Home name Hjem
TC2	Jolla (EN)	Settings		Home na… Hjem	Home name Hjem
TE1	Jolla (NO)	Settings		S?ker Hjem	Søker Hjem
TM1	Jolla (RU)	Settings		Hjem	Home name Hjem
SL1	Jolla (CZ)	Devices		Kapcsoló hozzáadása plug_product_names	Kapcsoló hozzáadása plug_product_names
TS2	Xperia Ray(2”)	Settings		Home name Hjem	Home name Hjem
SL2	Jolla (CZ)	Control		Stue -Vypnuto-	Stue AV

Some experiments – benchmarks (~80 apps)

A2DP Volume
aagtl
Aard
ACal
ADSDroid
agram
aGrep
aLogcat
aMAZEd
AppPushDrAngulo
Any Cut
AppTrack
arXiv mobile
Baby Care Timer
Battery
Bites
Blokish
Book
Catalogue
Campyre
CEToolbox
Addi
AutoAnswer
Datally
Ringdroid
Timer
Currency
Converter
Dalvik
Explorer
Dialer2
Handy Diary
Divide and
Conquer!
ES File
Explorer
Frozen
Bubble
Translate
Home
Manager
Hot Death
Jamendo
Music
KindMind
klaxon
Man Pages
Lock Pattern
Generator
Lolcat
Builder
K-9 Mail
Mileage
Mirrored
MNC Mobile
Multi SMS
Sender
MunchLife
My
Expenses
MyLock
Nectroid
NetCounter
Fake Dawn
aLogcat
SwiFTP
Quick Settings
PasswordMaker
Password
Manager
Random Music
Player
Pocket Talk
Rot13
Sanity
Soundboard
Text Edit
Tippy
Tomdroid
URforms
Database
Wamba
PhotoStream
WeightChart
Wikipedia
WordPress
WorldClock
Yet Another
Auto
Brightness
ZooBorns

Some findings

Readability, Style:

- Attention: for uninstalling Sanity you must revoke those privileges (otherwise the uninstall processo will fail).
- getUseWebViewBackgroundForOverscrollBackground : false
- If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.
- I/cr_LibraryLoader(3840): Expected native library version number "51.0.2704.90", actual native library version number "51.0.2704.90“
- Automatically remember last used media volume on disconnect
- The app needs access to your contects to tell who messages are from when reading messages.

Offensive:

- Error: invalid name!

Translations:

- Naujovinkite, kad vertimas neprisijungus būtų geresnis

“Conclusions” and further work

- The method is similar to an automated static analysis.
- Provided text presentation and content defects categories and defect types.
- Identification of the troubling user interface locations.
- The main drawback is long testing time (for images).
- What is the optimal subset of mobile devices for testing?

- Apps actually do have declared defects and it is possible to find them:
  - Apps translated automatically.
  - Partially translated apps.
  - Spelling.
  - Hard to understand texts.

Questions

Any bugs in presetation?