Data-Driven Requirements Engineering: The Way Ahead

Xavier Franch

Universitat Politècnica de Catalunya (UPC-BarcelonaTech)
Keynote at ENASE'20, May
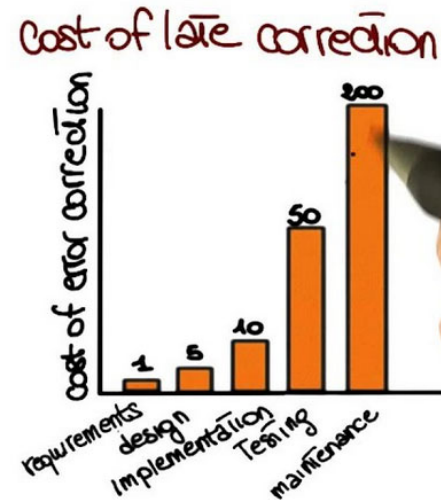**(virtually) Prague, Czech Republic**

ENASE 2020

15th International Conference on
Evaluation of Novel Approaches to Software Engineering
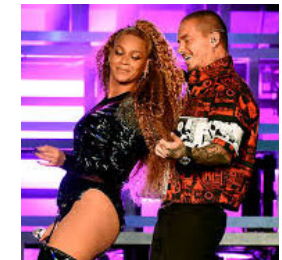
ONLINE STREAMING | 5 - 6 MAY, 2020

RE (still) central in SE

Prof. Ricardo Valerdi
rvalerdi@arizona.edu
University of Arizona & SpaceX

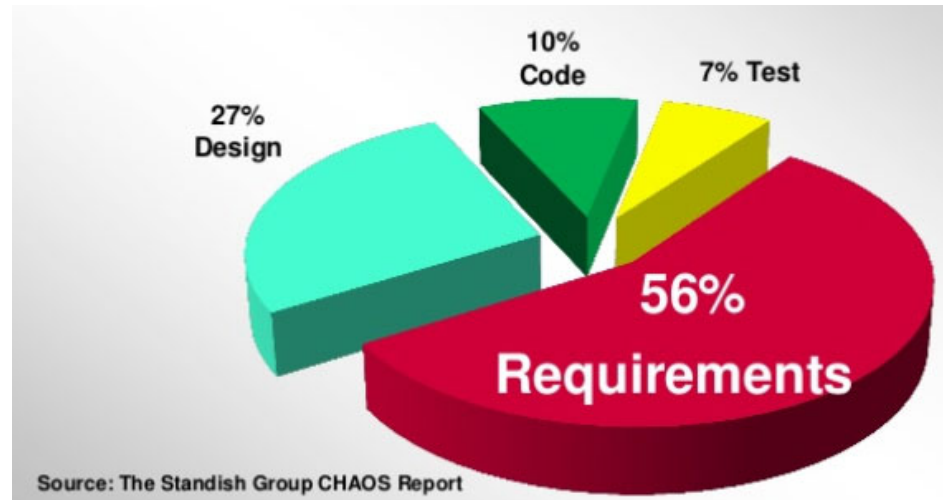## Take-Aways

1. Cost ≈ f(Effort) ≈ f(Size) ≈ f(Complexity)
2. Requirements understanding and "ilities" are the most influential on cost

2

RE (still) a problem in SE

Causes of failure

10% Code

7% Test

27% Design

56% Requirements

Source: The Standish Group CHAOS Report

THE STANDISH GROUP

PMI's
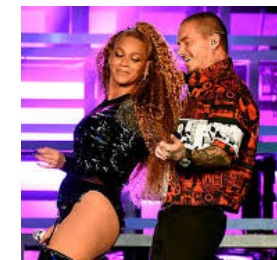PULSE of the PROFESSION®

Q: Of the projects started in your organization in the past 12 months that were deemed failures, what were the primary causes of those failures? (Select up to 3)
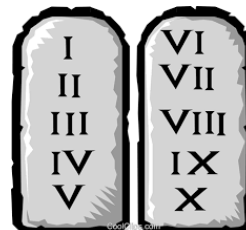
Global Total

Change in organization's priorities — 39%

Change in project objectives — 37%

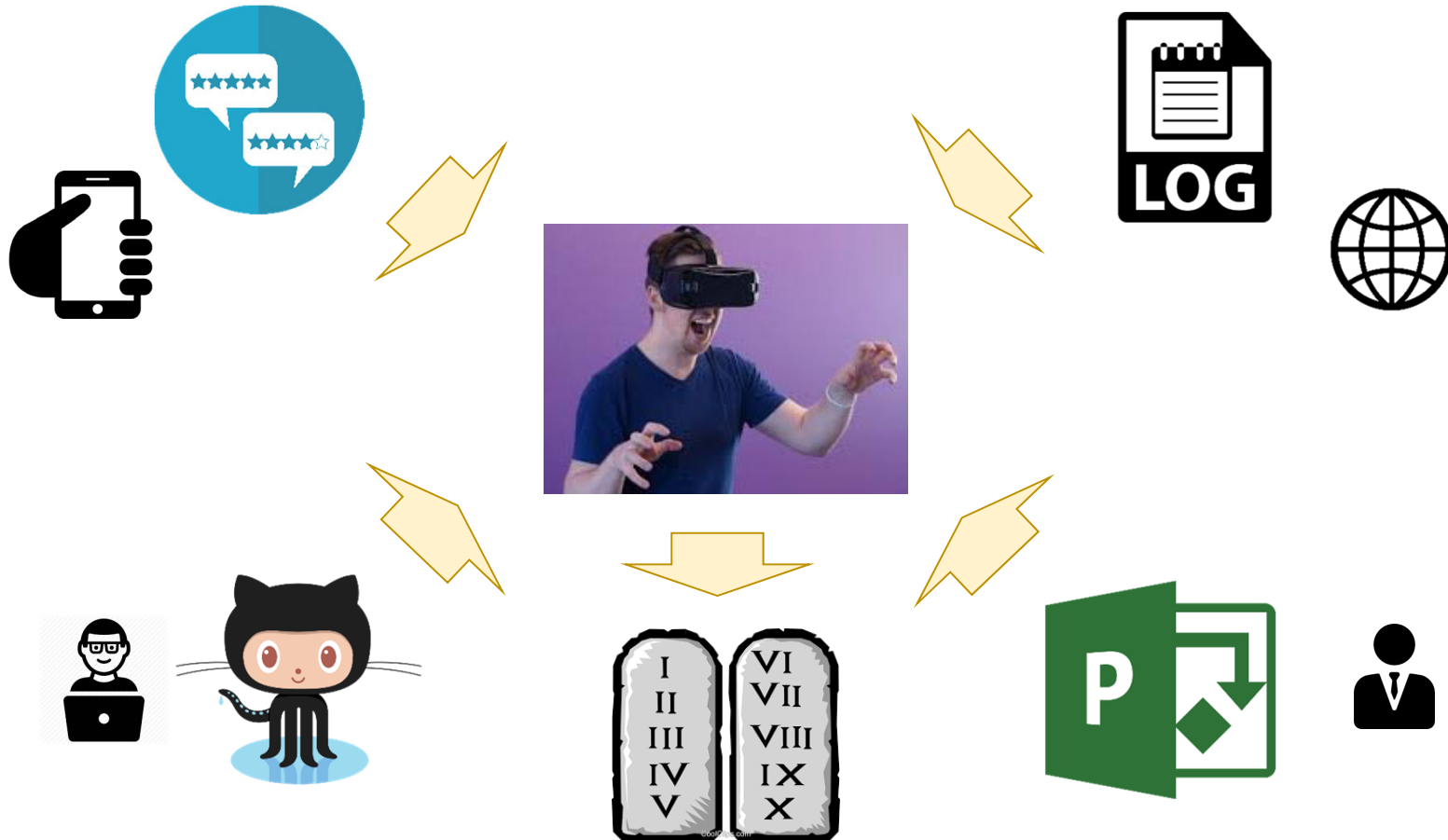Inaccurate requirements gathering — 35%

The question then is...

How to ensure that a system is delivering the right value to its stakeholders?
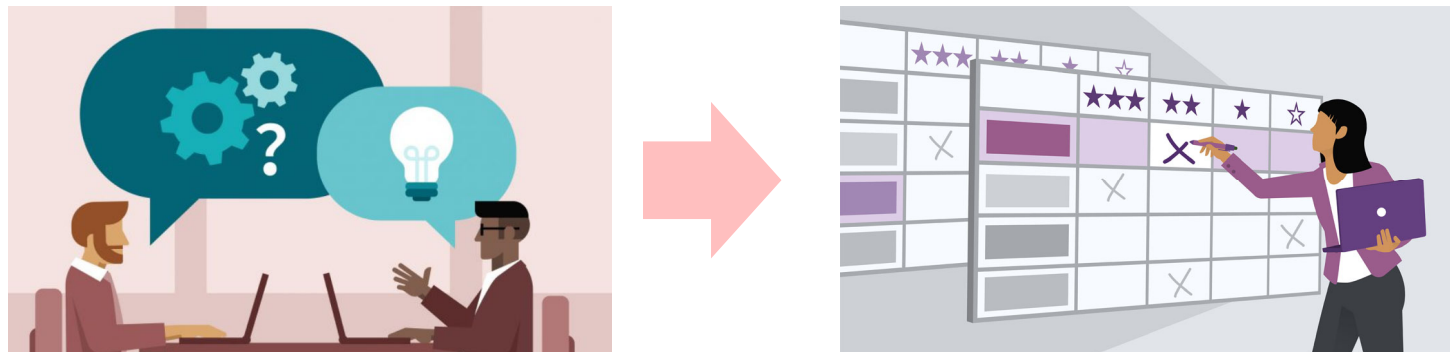
From traditional RE...

...to data-driven RE

Data-driven RE

Xavier Franch

Journals & Magazines > IEEE Software > Volume: 33 Issue: 1 ❓

Toward Data-Driven Requirements Engineering

4 Author(s)   Walid Maalej ; Maleknaz Nayebi ; Timo Johann ; Guenther Ruhe   **View All Authors**

"requirements engineering by the masses and for the masses"

Data-driven requirements engineering: an update

Full Text: 📄PDF 🛒 Get this Article

Authors: Walid Maalej   University of Hamburg
            Maleknaz Nayebi   Ecole Polytechnique Montreal
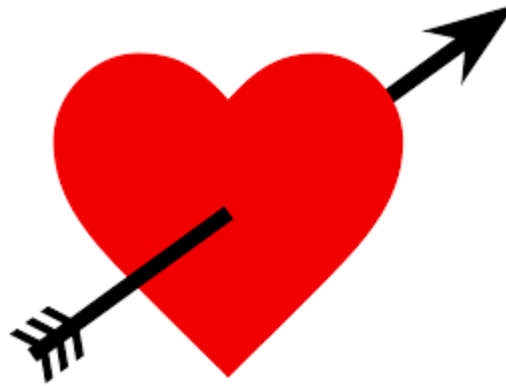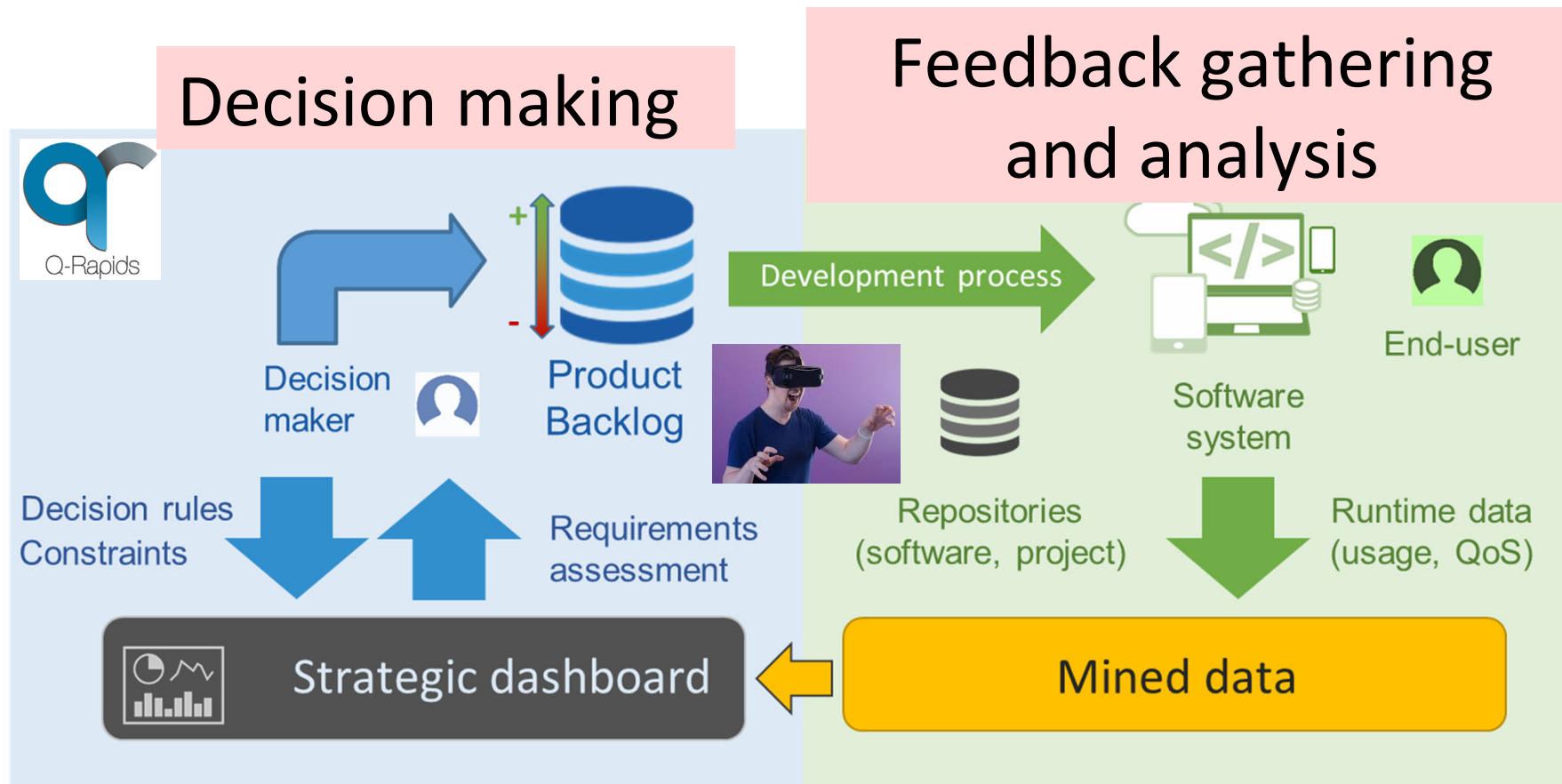   Guenther Ruhe   University of Calgary

Published in:
· Proceeding
ICSE-SEIP '19 Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice
Pages 289-290

Montreal, Quebec, Canada — May 27 - 27, 2019

A word of caution

The data-driven RE cycle

Decision making

Feedback gathering and analysis

Q-Rapids

Decision maker

Product Backlog

Development process

End-user

Software system

Decision rules Constraints

Requirements assessment

Repositories (software, project)

Runtime data (usage, QoS)

Strategic dashboard

Mined data

International Working Conference on Requirements Engineering: Foundation for Software Quality

REFSQ 2017: Requirements Engineering: Foundation for Software Quality pp 167-173 | Cite as

How Can Quality Awareness Support Rapid Software Development? – A Research Preview

Authors   Authors and affiliations

Liliana Guzmán ✉, Marc Oriol, Pilar Rodríguez, Xavier Franch, Andreas Jedlitschka, Markku Oivo

EXPLICIT FEEDBACK
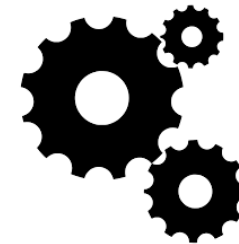
Explicit Feedback

- Feedback provided by the user at her will
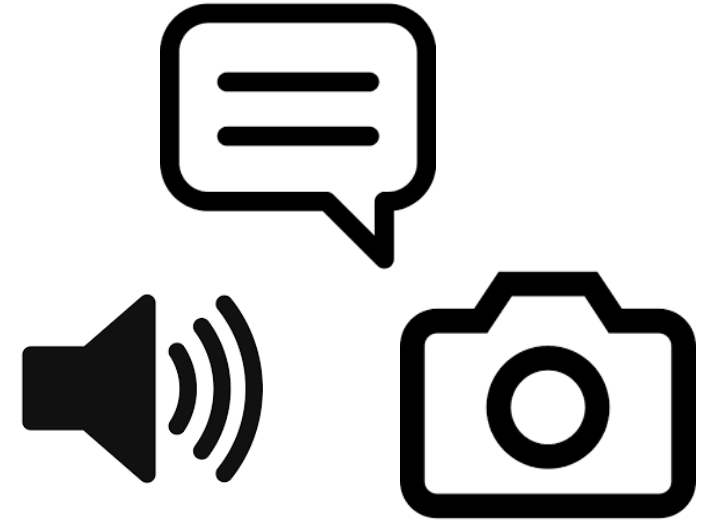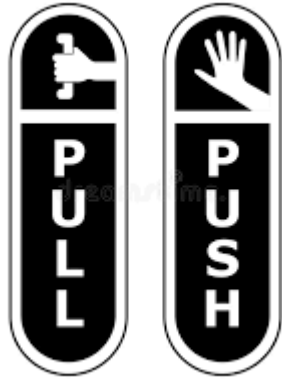- Main phases:

gathering

analysing

acting upon

Gathering explicit feedback

Explicit feedback analysis

- Preprocessing

- Categorization

- Sentiment analysis

- Topic modeling

Analysis: Preprocessing

- From text to lexical/syntactical units

  - Tokenization

  - Stemming / Lemmatization

  - Phrasing

    - Part-of-speech tagging

"I have a problem when saving the document, please check it"

I/**PRP** have/**VBP** a/**DT** problem/**NN** when/**WRB** saving/**VBG**
the/**DT** document/**NN** ,/, please/**VBP** check/**VB** it/**PRP**

Analysis: Categorization

- Bug report 🪲 or feature request ✋

Category	Definition
Feature shortcoming	Unsatisfying aspect of an existing feature.
Feature strength	Satisfying aspect of an existing feature.
Feature request	Request for a new feature.
Bug report	Report of an error, flaw, failure or fault.
Usage scenario	A way to use the software (e.g., recommended way, workaround).
Hardware constraint	Hardware needed to run the software.
Software constraint	Software needed to run the software.
General praise	General appreciation of the software focusing on the whole software system.
General complaint	General dissatisfaction of the software focusing on the whole software system.
Advertisement	Promotion of or suggestion to buy the software.
Dissuasion	Advise against the acquisition of the software.
Question	Question directly related to the software.
How to	Explanation to other users how to use the software.
Feature information	Description of a specific feature without any objective evaluation.
Software price	Discussion of the price of the software.
Compliance issue	Dispute over certain terms of agreement or regulations.
Software extension	Description of (planned) extensions of the software.
Other product	Reference to another software product.
Service	Comment on the service provided by the software.
Social interaction	Description of social/personal issues that arise from using the software (i.e., a software feature).
Content related	Comment about content that was created or is available through the software.
Job advertisement	Advertisement of a job available in the company developing the software.
Noise	Tweet not written in English or containing too many illegible symbols to be understandable.
Unclear	Tweet written in English, but the meaning of the tweet is ambiguous or unclear.
Unrelated	Tweet not related to the specific software at all.
Other	Tweet relevant for the study, but not covered by existing categories.

Conferences > 2016 IEEE 24th International ... ?

A Needle in a Haystack: What Do Twitter Users Say about Software?

3 Author(s)   Emitza Guzman ; Rana Alkadhi ; Norbert Seyff   View All Authors

15

Analysis: Sentiment analysis

Deciding if a piece of text expresses a particular affect or mood

+2 ⟶ [+2,-5] ⟵ -1 ⟶ -2 ⟵ -1 -1

had fun using it before but now it is really horrible :(help!!

Conferences > 2014 IEEE 22nd International ...

How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews

2 Author(s) Emitza Guzman ; Walid Maalej View All Authors

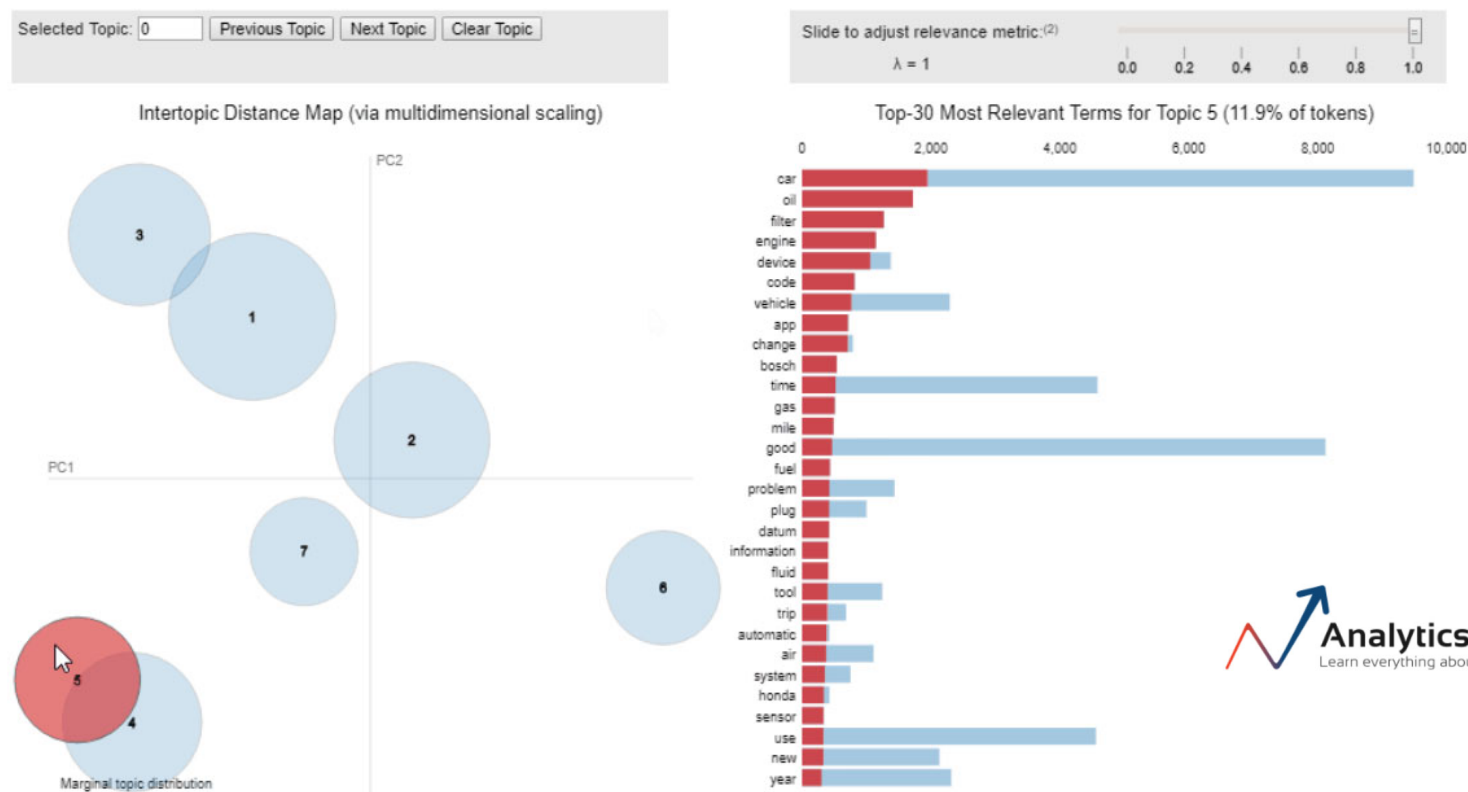uploading pictures with the app is so annoying! ⟶ [+1,-3]
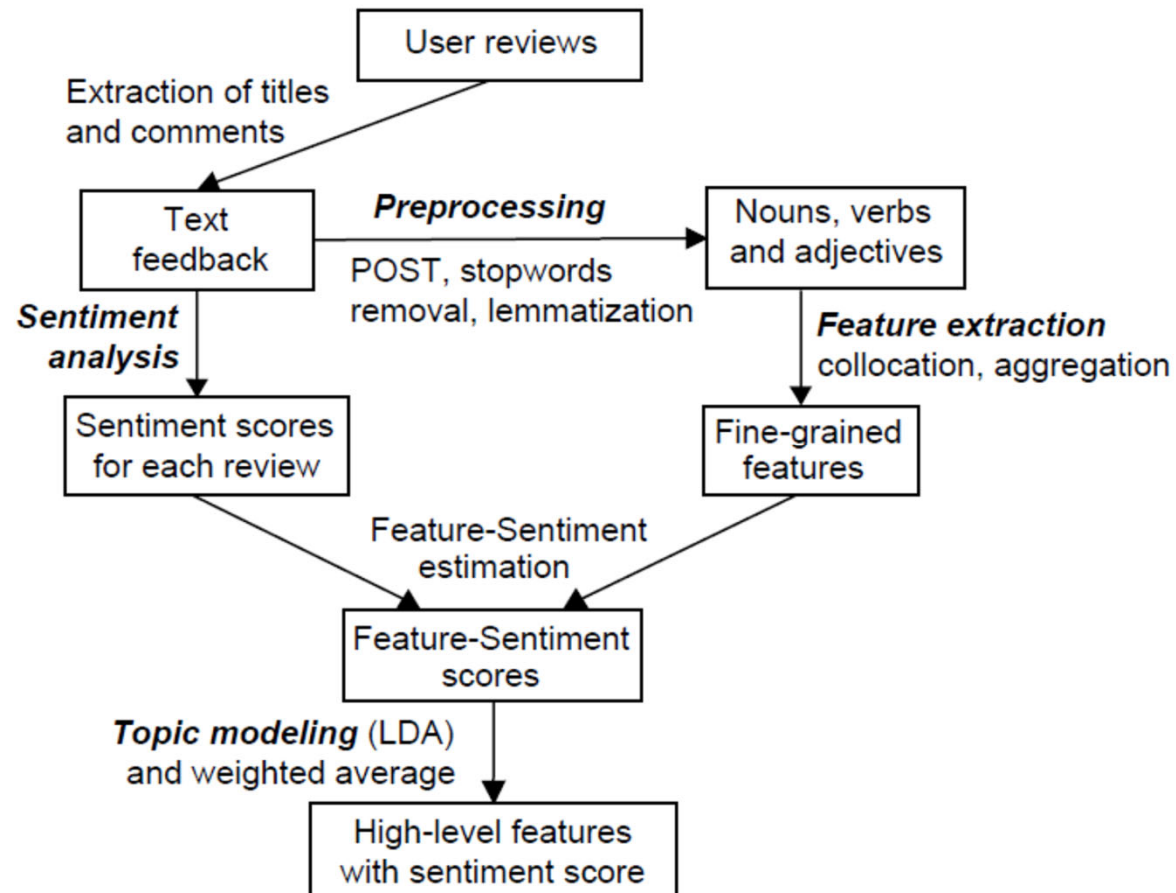
pleeeeeeease add an unlike button and I will love you forever!! ⟶ [+5,-1]

But of course, not easy...:

"Great, I like this feature that gives me this lovely headache"

Analysis: Topic modeling

- Identifying topics that best describes a corpus
  - Each topic described by a distribution of words
- Most popular algorithm: Latent Dirichlet Allocation

Putting all pieces together
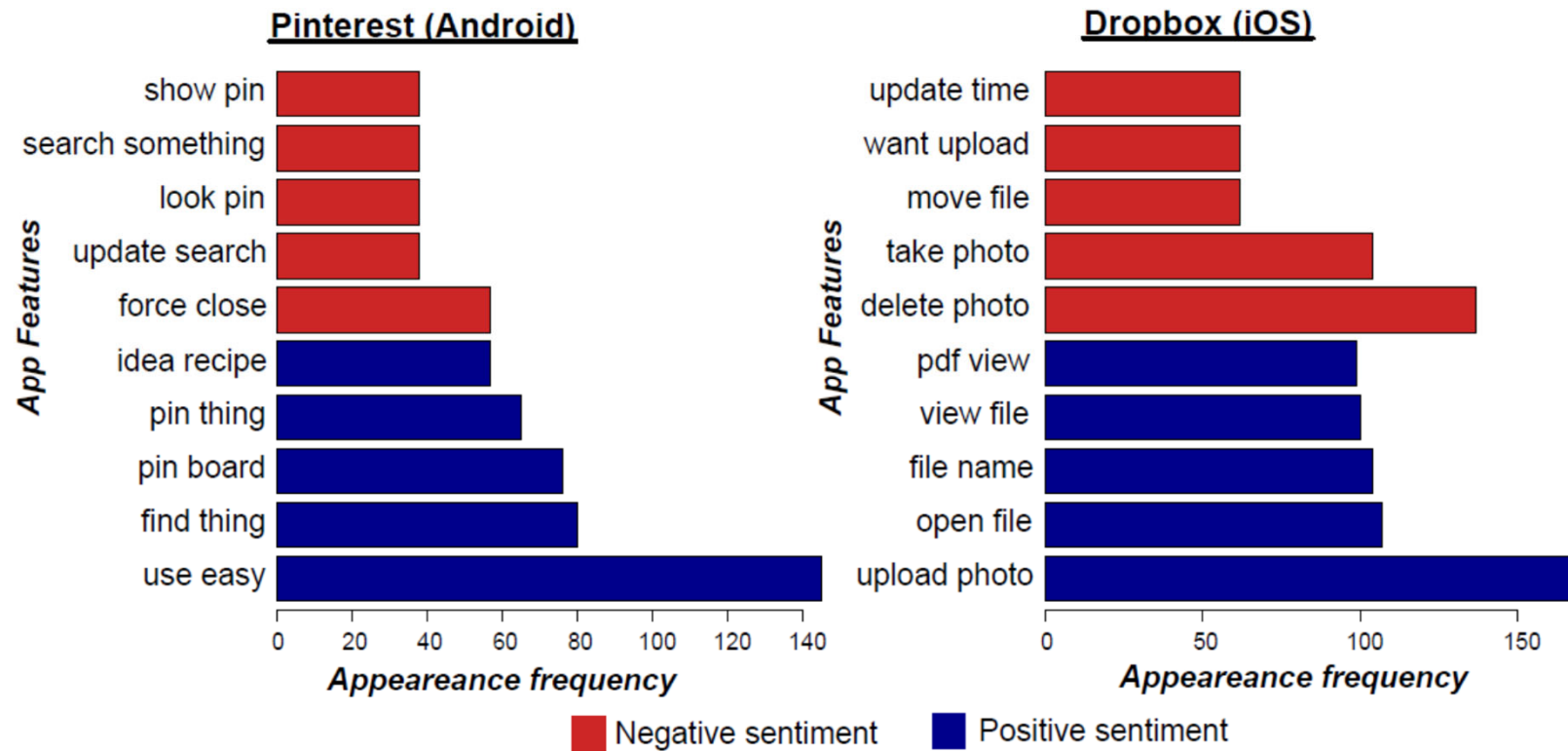
Pinterest (Android) / Dropbox (iOS) — Appearance frequency by App Features, Negative sentiment (red) and Positive sentiment (blue)

Conferences > 2014 IEEE 22nd International ...

How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews

2 Author(s)   Emitza Guzman ; Walid Maalej   View All Authors

Putting all pieces together

Extraction of titles and comments

User reviews

Text feedback

Preprocessing
POST, stopwords removal, lemmatization

Nouns, verbs and adjectives

Sentiment analysis

Feature extraction
collocation, aggregation

Sentiment scores for each review

Fine-grained features

Feature-Sentiment estimation

Feature-Sentiment scores

Topic modeling (LDA) and weighted average

High-level features with sentiment score

Conferences > 2014 IEEE 22nd International ...

How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews

2 Author(s) Emitza Guzman ; Walid Maalej View All Authors

Challenges with explicit feedback

MOTIVATION

PRIVACY

RELIABILITY

Towards Understanding and Detecting Fake Reviews in App Stores

Daniel Martens · Walid Maalej

EMPIRICAL SOFTWARE ENGINEERING

IMPLICIT FEEDBACK

Monitoring QoS
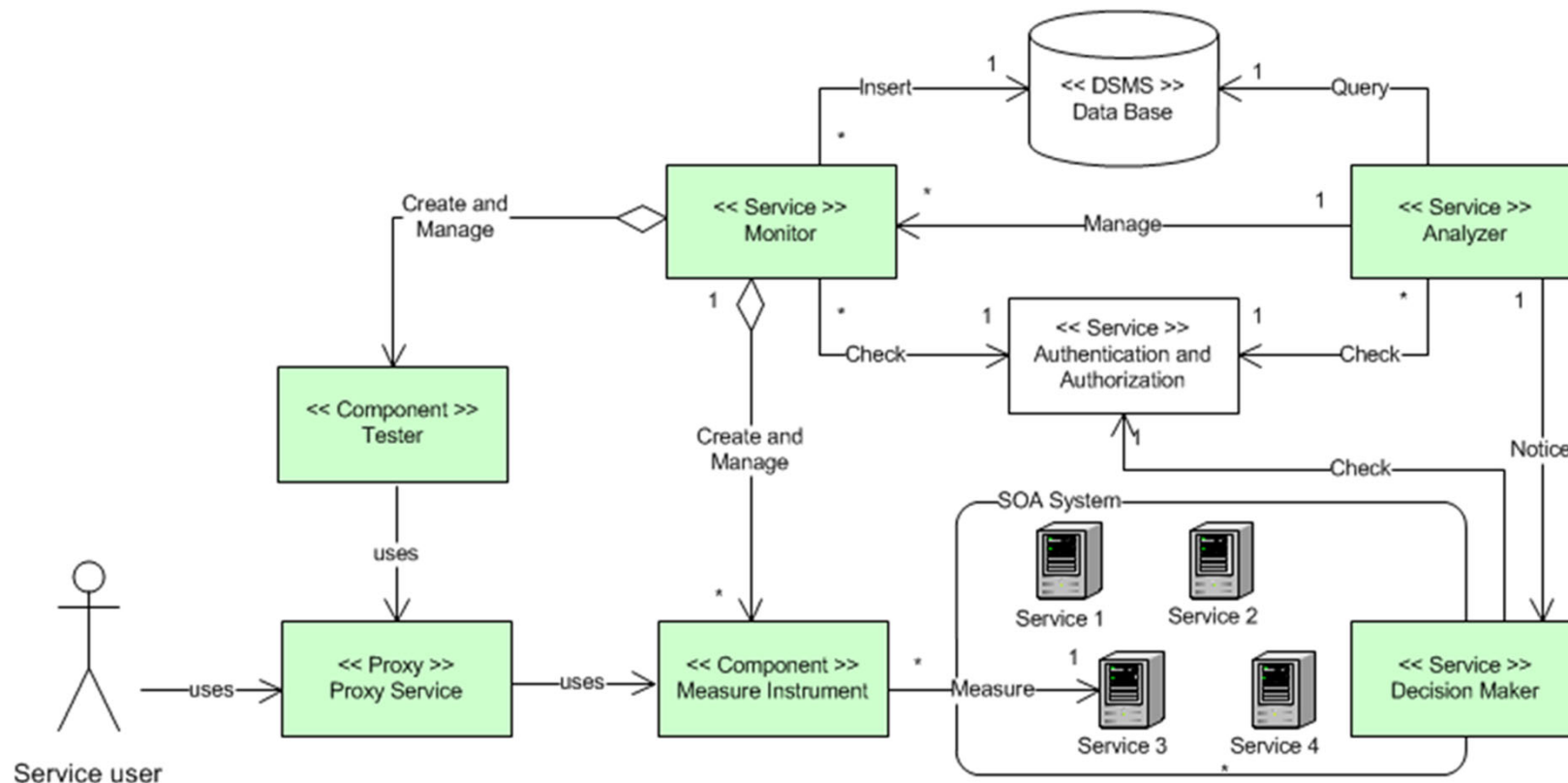
Monitoring the service-based system lifecycle
with SALMon

Marc Oriol ⌂ ✉, Xavier Franch ✉, Jordi Marco ✉

Usage logs

- Information about usage of the system

timestamp	event Type	element type	element ID	text	element value	URL
01/10/2017 4:36:03	click	INPUT	meter reading			https://www.energiesparkonto.de/esk/content/startPage/?portal_id=co2online
01/10/2017 4:36:09	click	INPUT	enterMeterreading_save		Save meter reading	https://www.energiesparkonto.de/esk/content/startPage/?portal_id=co2online
01/10/2017 4:37:30	click	SELECT	view	Years \t Months \t Weeks \t Days\t Hours \t	year	https://www.energiesparkonto.de/esk/content/bereichePage/?bereich=strom
01/10/2017 4:37:32	click	OPTION		Months	month	https://www.energiesparkonto.de/esk/content/bereichePage/?bereich=strom
01/10/2017 4:38:21	click	SELECT	view	Years \t Months \t Weeks \t Days\t Hours \t	month	https://www.energiesparkonto.de/esk/content/bereichePage/?bereich=strom
01/10/2017 4:38:23	click	OPTION		Days	day	https://www.energiesparkonto.de/esk/content/bereichePage/?bereich=strom
01/10/2017 4:39:07	click	A	logout	Logout	undefined	https://www.energiesparkonto.de/esk/content/bereichePage/?bereich=strom

- What can be discovered?
  - Which functionalities are most used
  - Which navigational paths prevail (or not)
  - Which calls result often in error codes
  - …
  - Especially if QoS also available

Challenges with usage logs

- Data noise

- Incomplete data

- Concept of session

- Evolution

- Anonymization

- ...

Importance of context
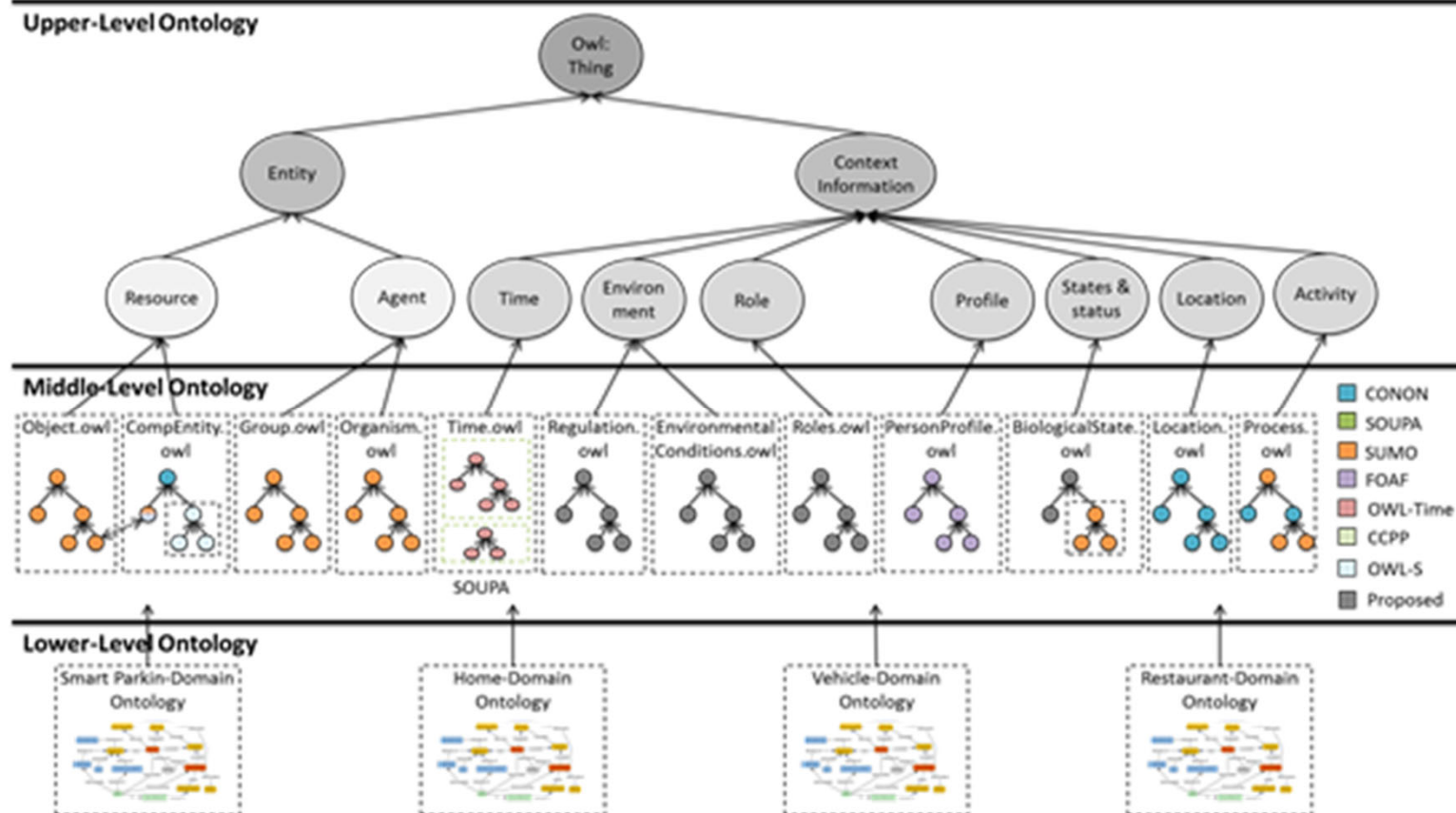
3LConOnt: a three-level ontology for context modelling
in context-aware computing

Authors   Authors and affiliations

Oscar Cabrera ✉ , Xavier Franch, Jordi Marco

DECISION MAKING

Keys to decision making

software analytics

release planning
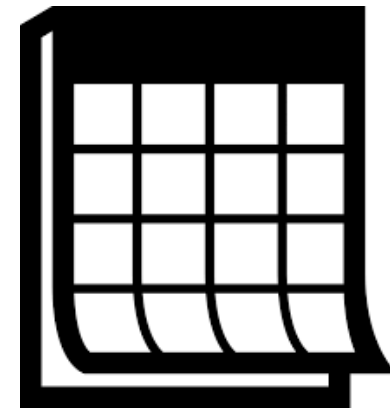
Code analytic tools
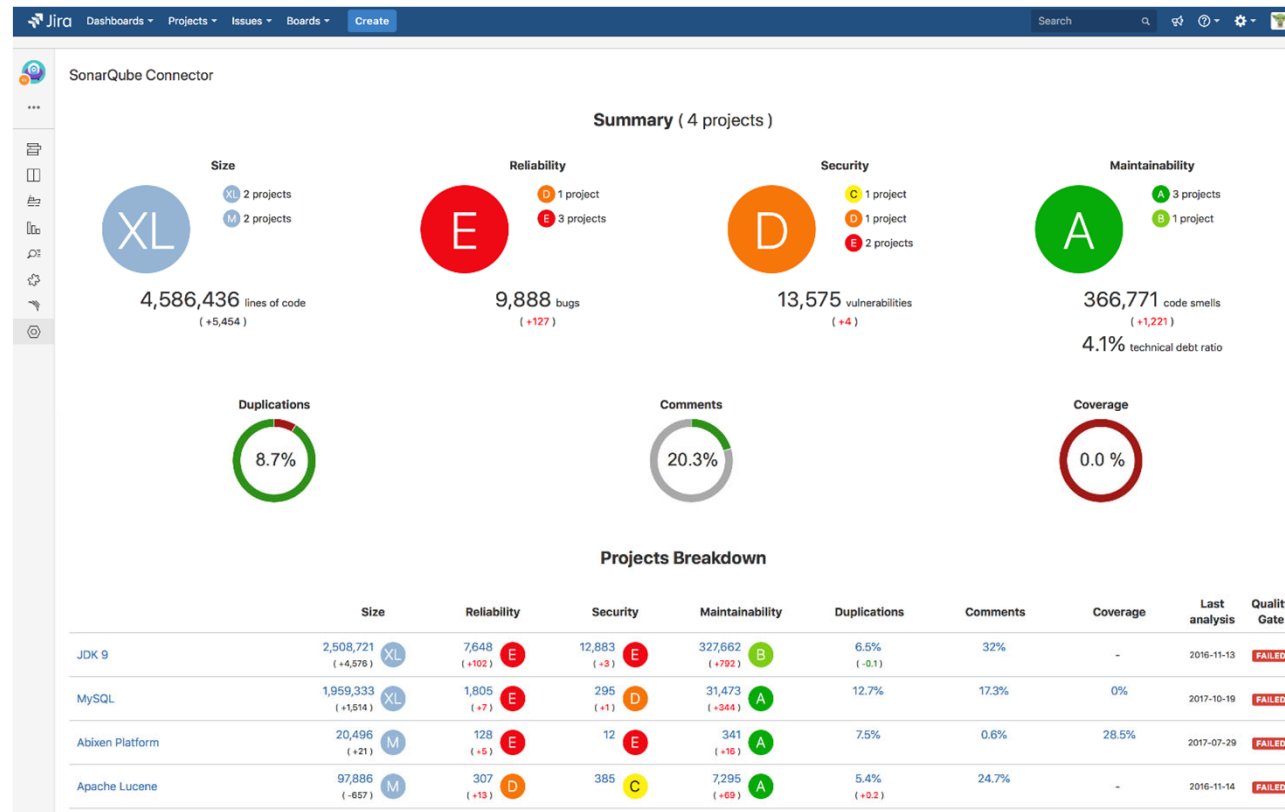
Valuable, but not strategic enough

Analytic dashboards

Elaborating indicators
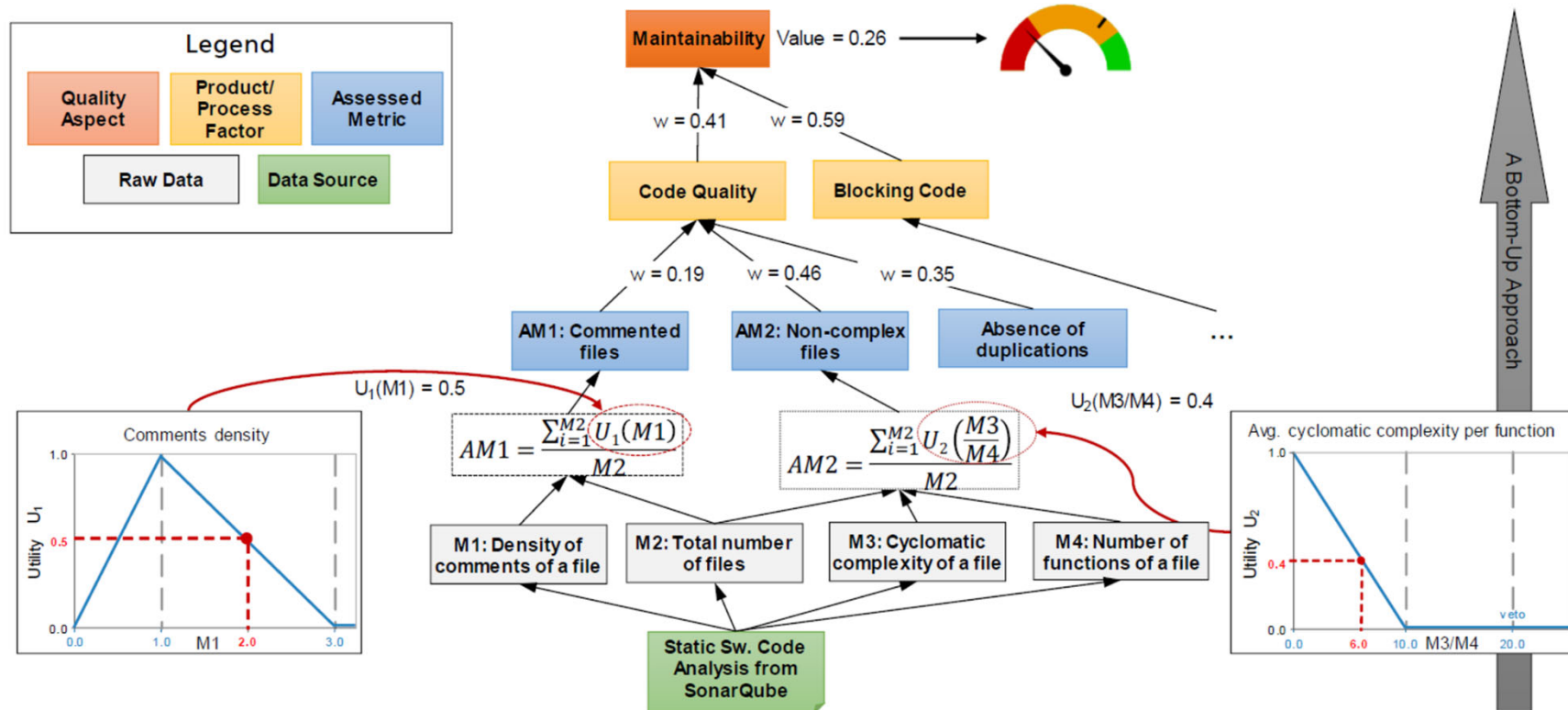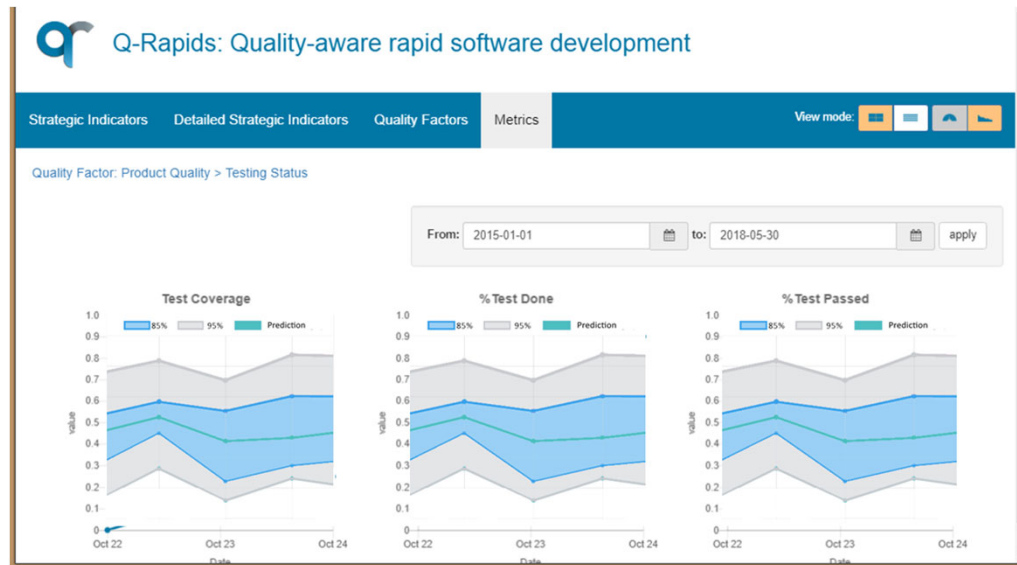
30

Techniques

prediction

what-if analysis
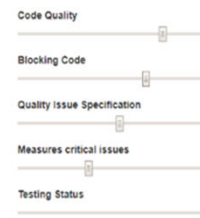
Release planning

- Closing the data-driven RE cycle

$F = \{f(1), ..., f(N)\}$

Set of requirements

$C = \{c(1), ..., c(M)\}$

Set of constraints

RP

maximise some utility
or objective function

$X = \{x(1), ..., x(N)\}$

Release plan
$x(j)$ = assigned release
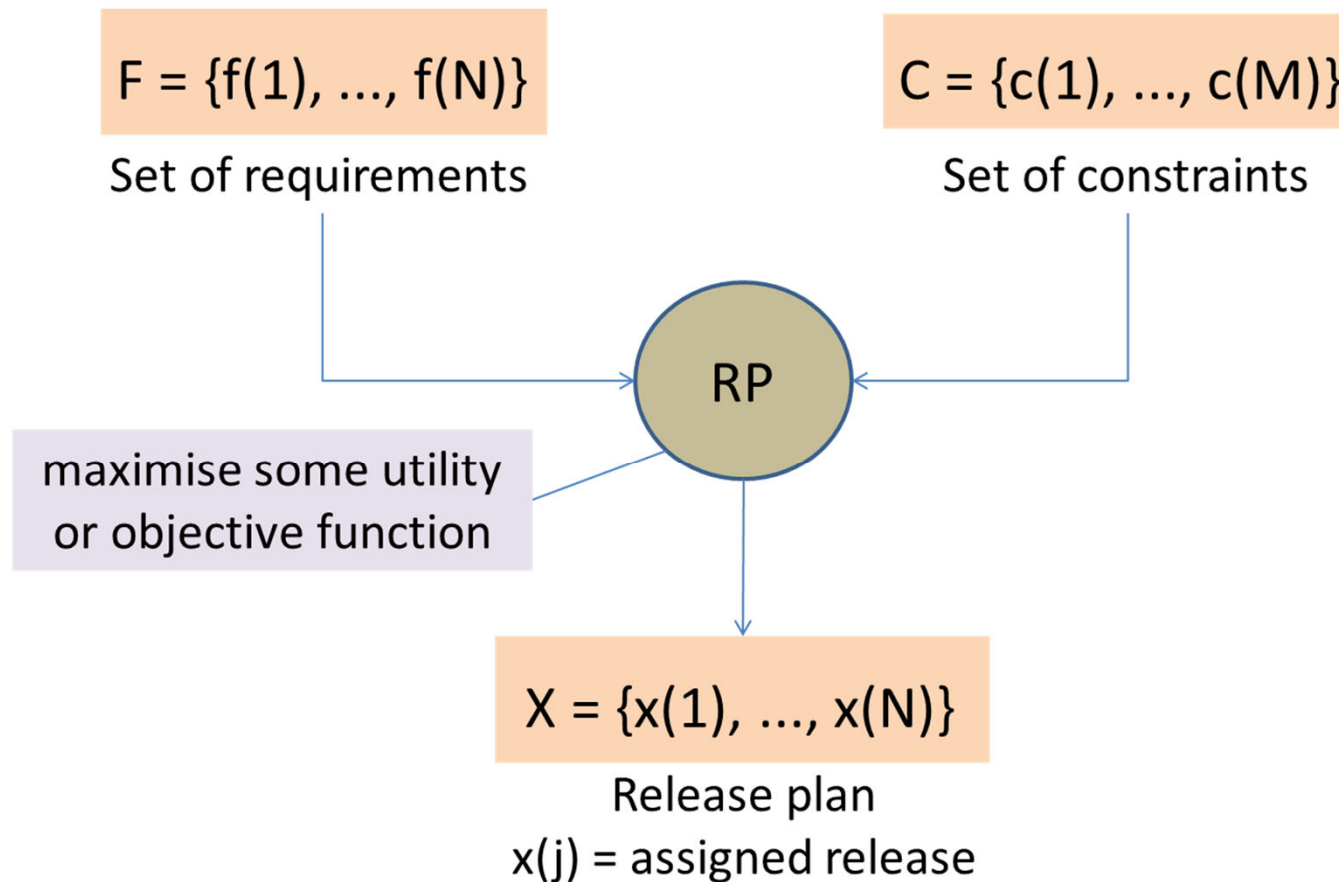
Release Planning in DDRE
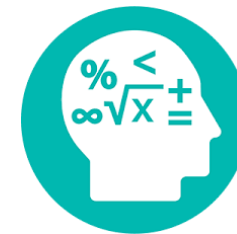
REACTIVE

PROACTIVE

Journals & Magazines > IEEE Software > Volume: 33 Issue: 1

Toward Data-Driven Requirements Engineering

4 Author(s)   Walid Maalej ; Maleknaz Nayebi ; Timo Johann ; Guenther Ruhe   **View All Authors**

Data-driven requirements engineering: an update

Authors: Walid Maalej, Maleknaz Nayebi, Guenther Ruhe   Authors Info & Affiliations

Publication: ICSE-SEIP '19: Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice • May 2019 • Pages 289–290 • https://doi.org/10.1109/ICSE-SEIP.2019.00041

33

Involving relevant stakeholders

Liquid democracy

- Stakeholders can evaluate requirements based on different interest dimensions
  - Most stakeholders might not be appropriate to evaluate all requirements accurately
  - Stakeholders can selectively delegate votes to experts

THE WAY AHEAD

The way ahead

- Six research challenges for data-driven RE

International Working Conference on Requirements Engineering: Foundation for Software Quality

REFSQ 2020: Requirements Engineering: Foundation for Software Quality pp 135-142 | Cite as

Towards Integrating Data-Driven Requirements Engineering into the Software Development Process: A Vision Paper

Authors   Authors and affiliations

Xavier Franch, Norbert Seyff, Marc Oriol ✉ , Samuel Fricker, Iris Groher, Michael Vierhauser, Manuel Wimmer

Conference paper
First Online: 18 March 2020

124
Downloads

Part of the Lecture Notes in Computer Science book series (LNCS, volume 12045)
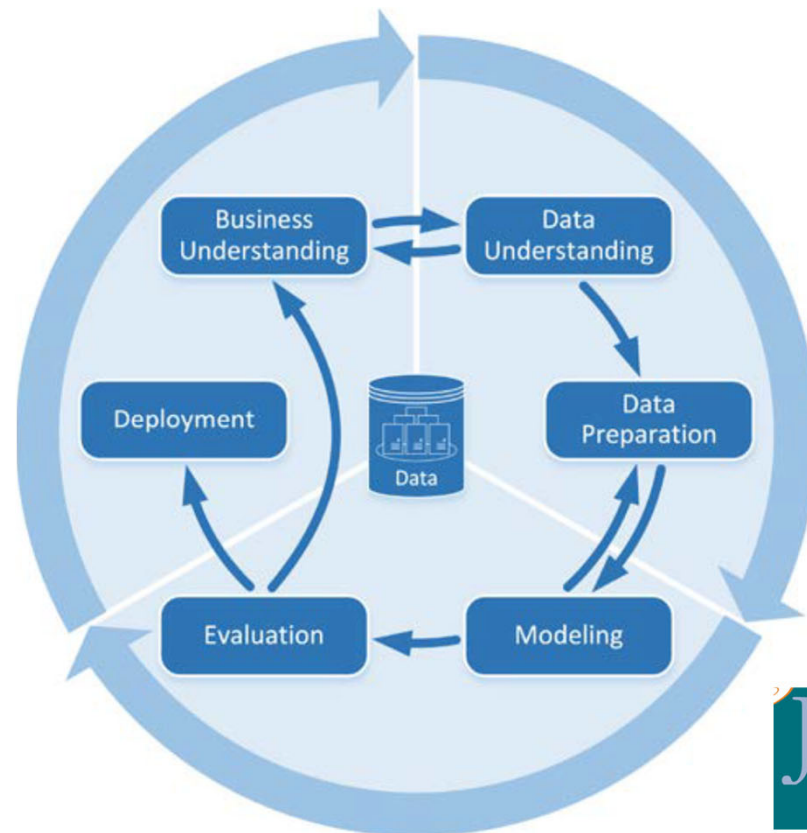
Challenge 1

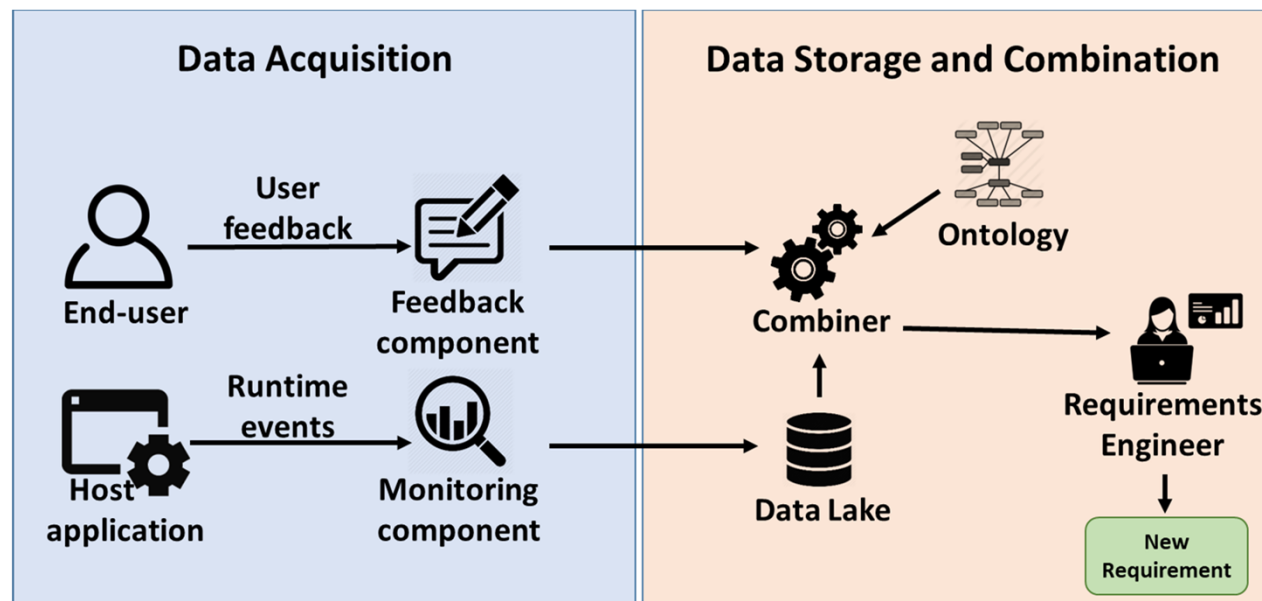Seamless integration into existing development processes and software systems
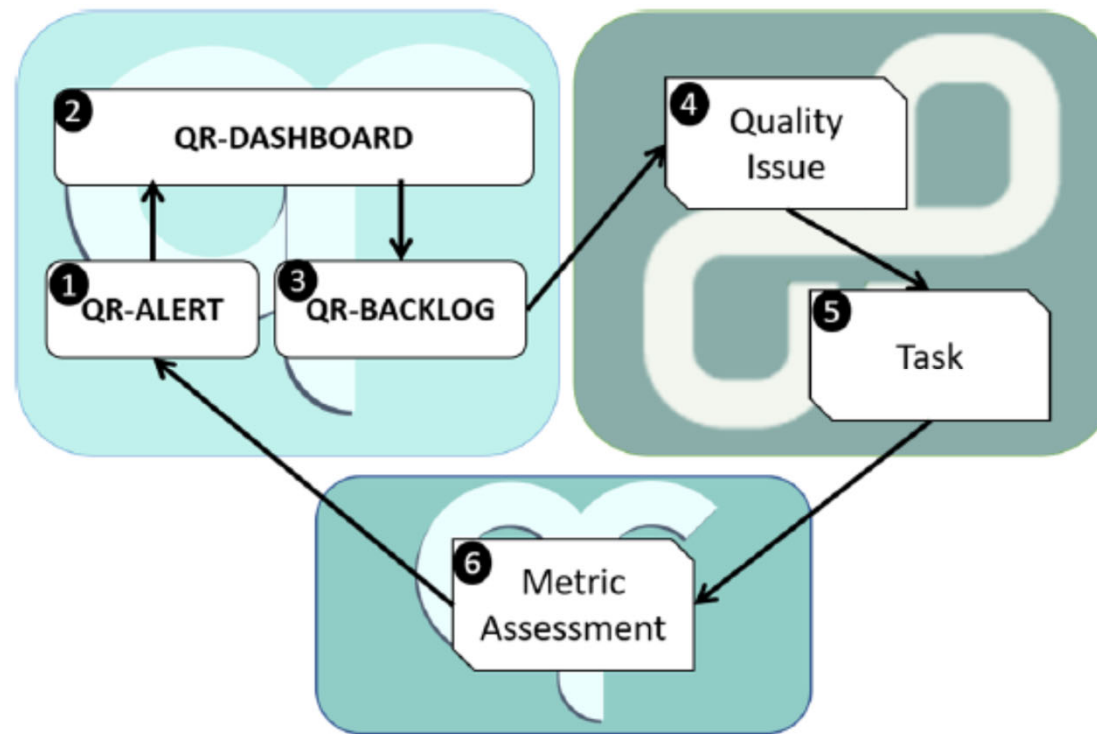
Challenge 1

Seamless integration into existing development processes and software systems

Challenge 2

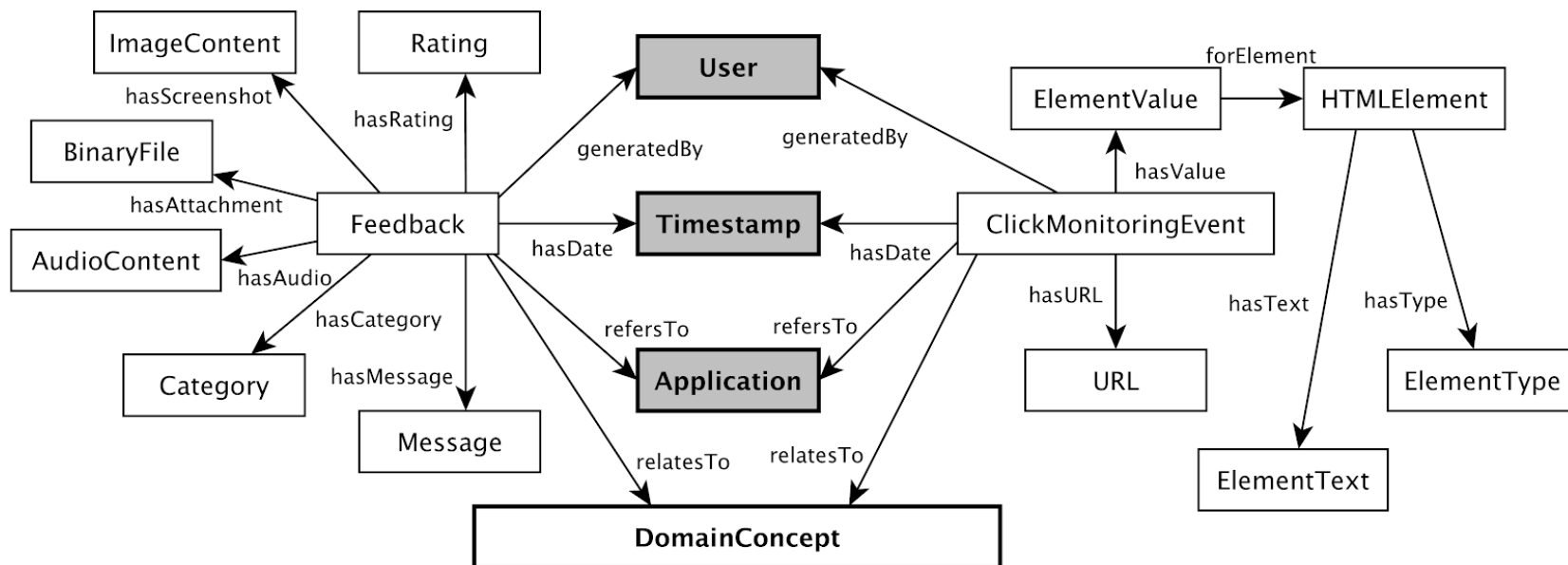Collection, processing, and integration of relevant heterogeneous information

FAME: Supporting Continuous Requirements Elicitation by Combining User Feedback and Monitoring

Challenge 2

Collection, processing, and integration of relevant heterogeneous information

Challenge 3

Context-awareness and adaptability
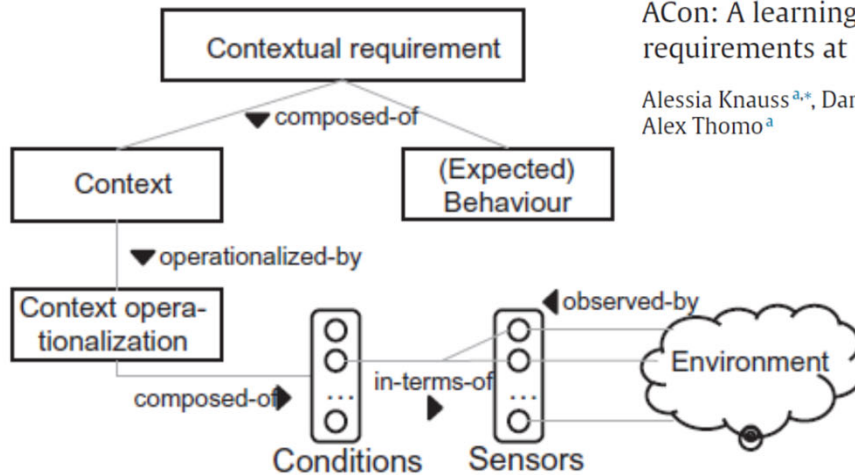
Information and Software Technology
Volume 105, January 2019, Pages 161-189

ELSEVIER

Adaptive monitoring: A systematic mapping

Edith Zavala, Xavier Franch, Jordi Marco

Self-adapting monitoring

ACon: A learning-based approach to deal with uncertainty in contextual requirements at runtime

Alessia Knauss[a,*], Daniela Damian[a], Xavier Franch[b], Angela Rook[a], Hausi A. Müller[a], Alex Thomo[a]

$cr_i \in CR$	Context (c_i)	Exp. behaviour (b_i)	Operationalization of context ($Oper(c_i)$)	Sensors involved ($vars(Oper(c_i))$)
cr_1	Driver is sleeping	Support lane keeping	$((BlinkOfEye \leq 0.9)$ and $(PositionOfHead \leq 0.85)$ and $(PositionOfHead \geq 0.75))$ or $((RightArmOffSteeringwheel = 1)$ and $(LeftArmOff\text{-}Steeringwheel = 1)$ and $(PositionOfHead > 0.95))$	BlinkOfEye, PositionOfHead, RightArmOffSteeringwheel, LeftArm-OffSteeringwheel, PositionOfHead

| Cases	Detection of uncertainty	Condition on context (c)	Condition on behaviour (b)
Case 1	No operationalized context.	$c \notin dom(Oper)$	
Case 2 a)	Sensor lost	$\exists e \in vars(Oper(c)): e \notin dom(Meas)$	
Case 2 b)	Sensor decalibrated	$\exists e \in vars(Oper(c)): outlier(e)$	
Case 2 c)	Sensor up	$\exists e \in vars(Oper(c)):$ Time $t-1: \neg correctSensor(e)$ Time $t: correctSensor(e)$	
Case 3	Violation	$Eval_{Context}(c) = true$	$Eval_{Behav}(b) = false$
Case 4	Potentially wrong context	$Eval_{Context}(c) = false$	$Eval_{Behav}(b) = true$
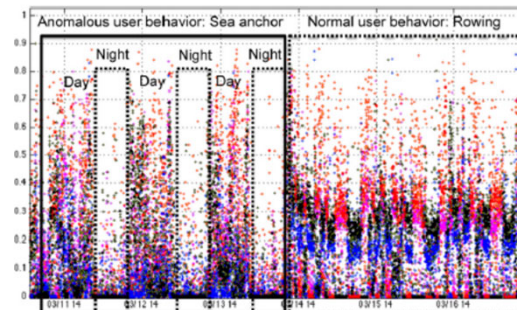
Challenge 4

Provision of actionable feedback
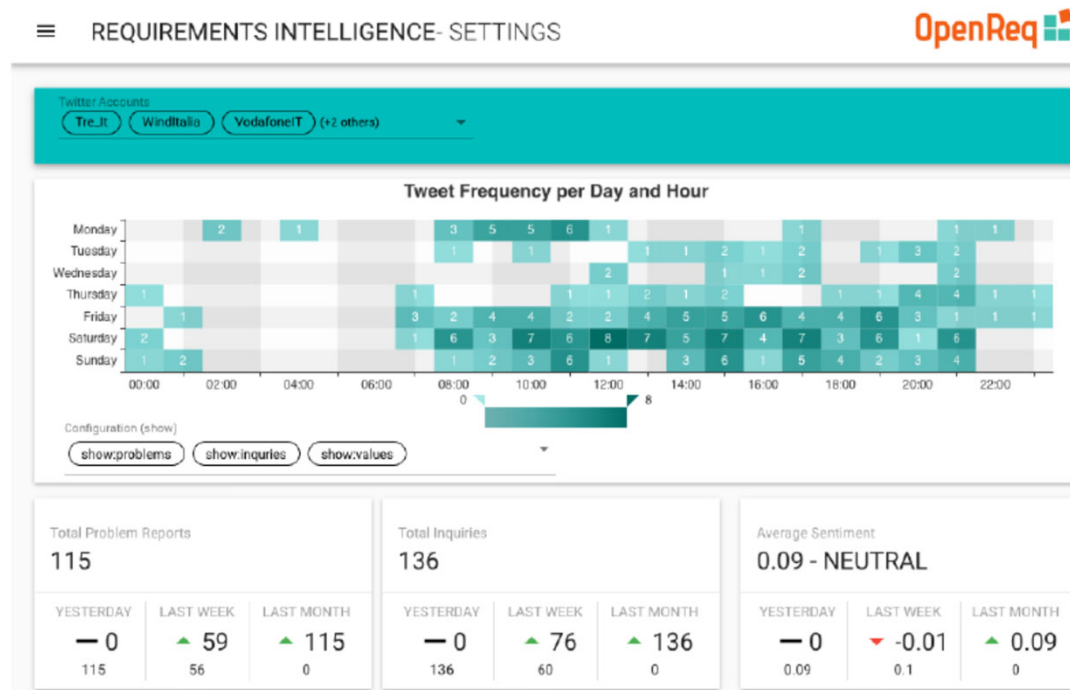
Challenge 5

Gaining users' trust

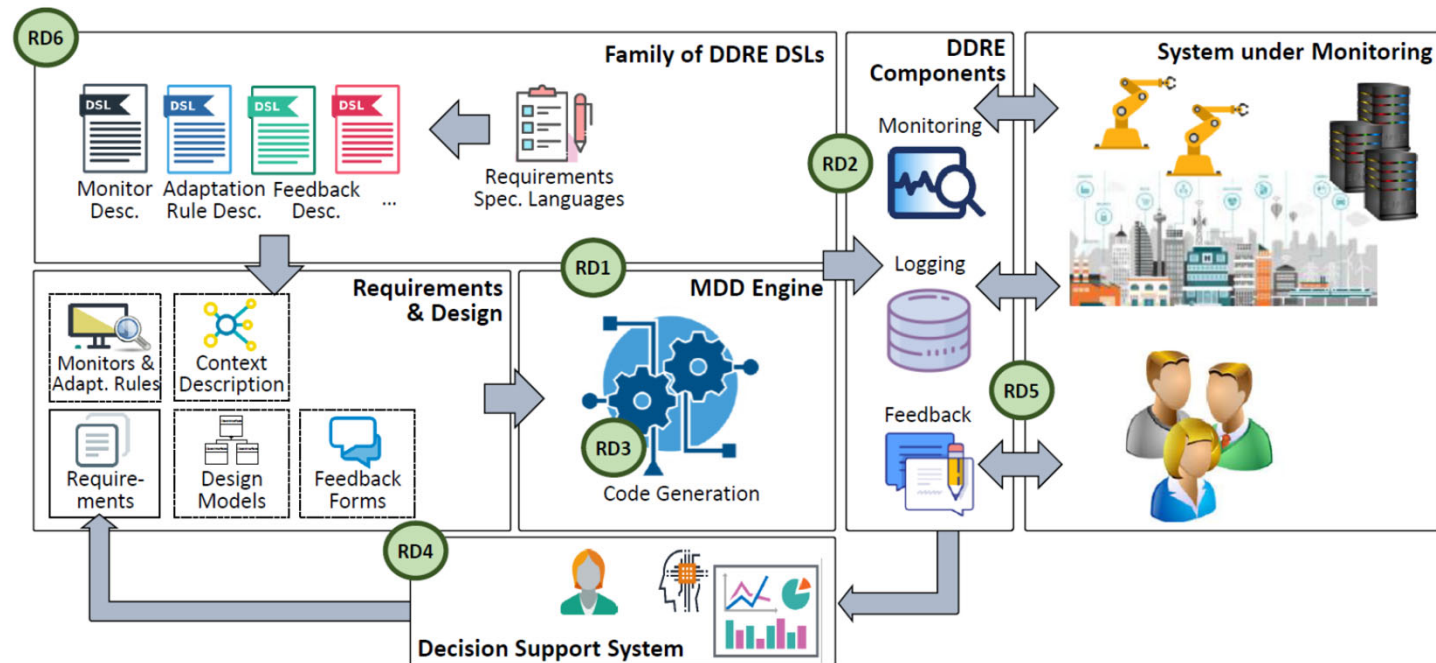Server (Federated anonymization service) | Server (Federated Application)
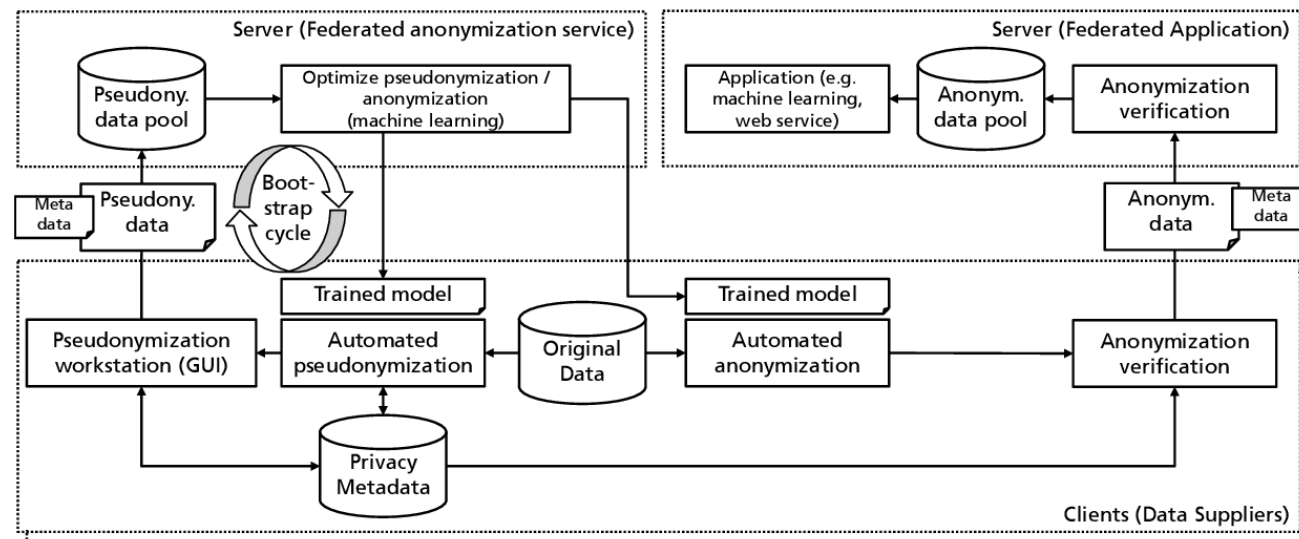
Published in ICEIS 2019

Towards a Privacy Compliant Cloud Architecture for Natural Language Processing Platforms

Matthias Blohm, Claudia Dukino, Maximilien Kintz, Monika Kochanowski, Falko Koetter, Thomas Renner

Challenge 6

Provision of value for the entire life-cycle

Challenges to adoption

- Organizational
  - Tailoring to the company
  - Integration with company WoW
  - Shared vocabulary

- Value
  - Informativeness
  - Transparency

- Technological
  - Simplify tool installation
  - Efficient tool configuration

Journals & Magazines > IEEE Access > Volume: 7 ❓

Continuously Assessing and Improving Software Quality With Software Analytics Tools: A Case Study
Publisher: IEEE

11 Author(s)   Silverio Martínez-Fernández ⓘ ; Anna Maria Vollmer ; Andreas Jedlitschka ; Xavier Franch ; Lidia López ...   **View All Authors**

Success factors

- Organizational
  - Incremental adoption
  - Monitor progress with strategic indicators
  - Involve experts

- Value
  - Transparency as a business value
  - Tailoring to different scopes

- Technological
  - Single access point to software quality related data

CONCLUSIONS

Data-driven RE

- Offers a great opportunity for delivering more business value to systems' stakeholders

- Some considerations

  - Not a hammer for every nail

  - Data-driven → needs data

    - Still traditional methods at least to start with

    - The role of traditional RE in the loop is worth considering

  - New, but maybe not as much

    - E.g., compare to experimentation (from lean startup)

      - A/B tests, fake door tests, ...

ELSEVIER

Information and Software Technology
Volume 77, September 2016, Pages 80-91

Raising the odds of success: the current state of experimentation in product development

Eveliina Lindgren [a] [✉], Jürgen Münch [a] [b] [✉]

Images

All other images labelled for reuse

https://dumielauxepices.net/sites/default/files/panels-clipart-focus-group-discussion-706060-4505275.png

https://planningshopintl.com/solution/ethnography/

https://www.lrmhc.org/2017-community-health-needs-survey/

http://es.coolclips.com/m/vector/reli0055/Diez-Mandamientos

https://www.flaticon.com/free-icon/log-file-format_28822

https://www.flaticon.com/free-icon/manager_115893

https://www.iconfinder.com/icons/3356606/computer_geek_glasses_nerd_icon

https://www.pinterest.es/pin/802625964810288333

https://commons.wikimedia.org/wiki/File:Globe_icon.svg

https://www.glocksoft.com/email-marketing-software/web-analytics-tools

https://www.onlinewebfonts.com/icon/488455

https://www.123rf.com/photo_100296089_stock-vector-pray-or-hands-together-in-prayer-line-art-vector-icon-for-religious-apps-and-websites.html

http://blog.netumo.com/monitoring-the-internet-of-things-iot

https://www.onlinefreecourse.net/lynda/requirements-elicitation-for-business-analysts-interviews-lynda-free-download

https://www.lynda.com/Project-Management-tutorials/Applying-Agile-MoSCoW-Prioritization/718626-2.html

https://www.pinterest.es/pin/97460779420537627/

https://www.sandiego.gov/pad/what-we-do

https://www.nqbe.com.au/single-post/2018/08/13/Proactive-vs-Reactive---Prevention-vs-Cure

Data-Driven Requirements Engineering: The Way Ahead

Xavier Franch

Universitat Politècnica de Catalunya (UPC-BarcelonaTech)
Keynote at ENASE'20, May
(virtually) Prague, Czech Republic

ENASE 2020
15th International Conference on
Evaluation of Novel Approaches to Software Engineering
ONLINE STREAMING | 5 - 6 MAY, 2020