Tool support for architectural decision making in large software-intensive agile projects

Manoj Bhat, Klym Shumaiev, Florian Matthes

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Motivation

- Software architecture is considered as a set of architectural design decisions.

- Design decisions link stakeholders' concerns and the architectural elements.

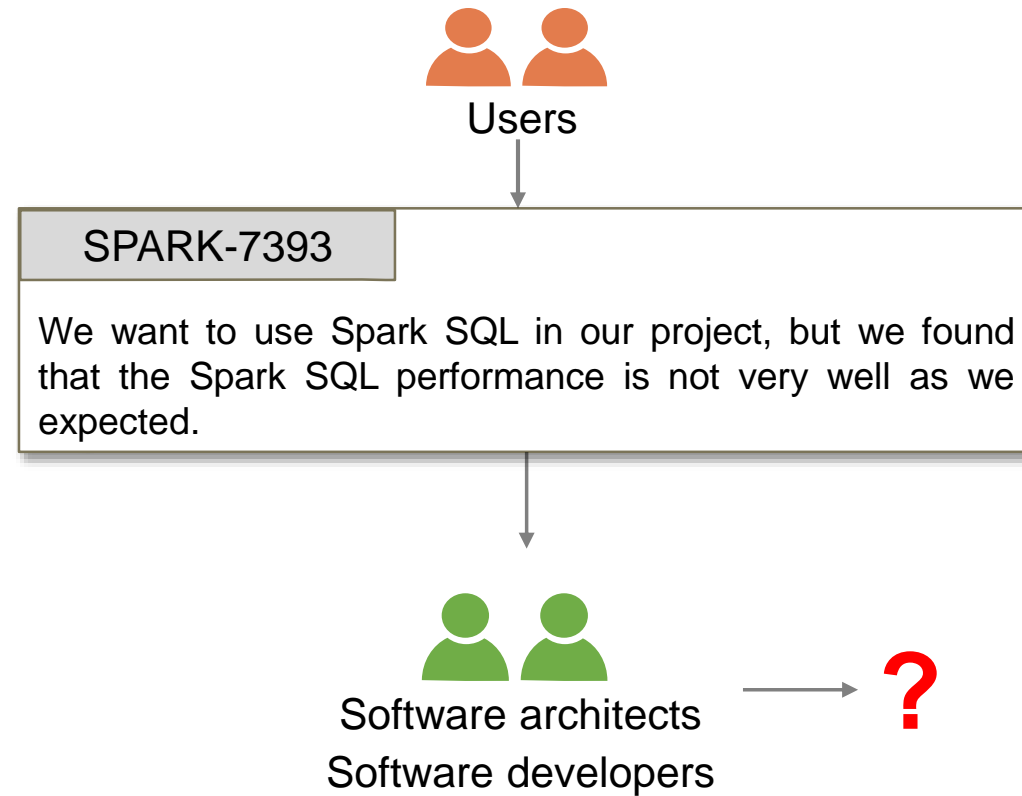- Design decisions are *not* explicitly documented which leads to knowledge evaporation.

Goal

- Framework for managing design decisions in **large software-intensive agile projects**.

- Support architects and developers to learn from past design decisions to make **sustainable informed decisions** in their current projects.

- Provide a perspective on **Who did What, Why and When** in the software architecture decision-making context.
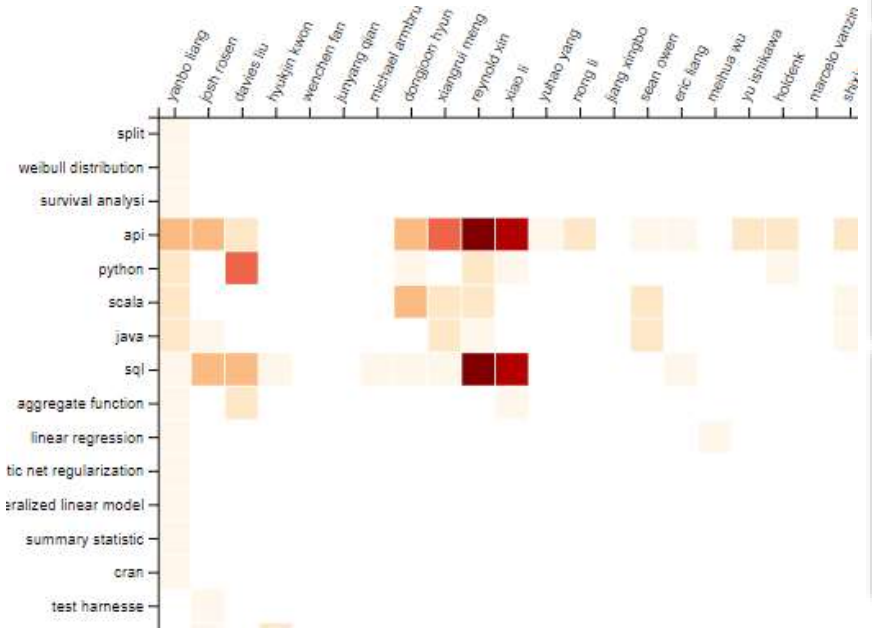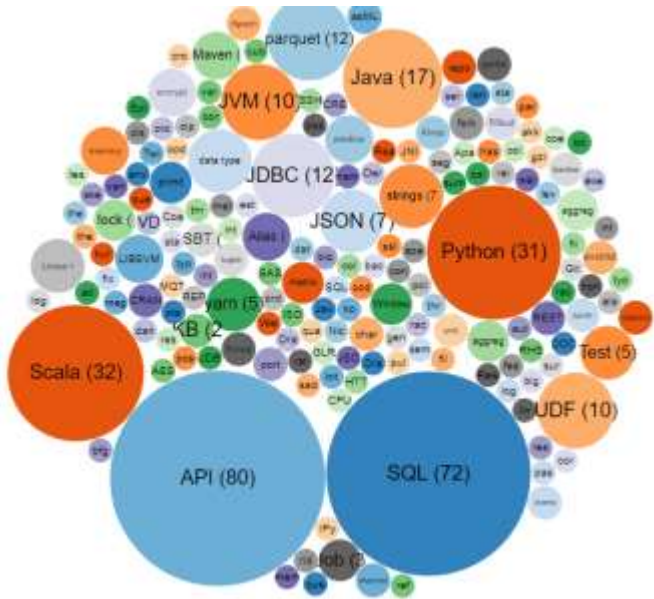
Example scenario (1)

Apache Spark

- Issues captured since early 2014
- Versions from 0.9.0 to 2.1.0
- # contributors - **1,146**
- # issues - **21,660**

Users

SPARK-7393

We want to use Spark SQL in our project, but we found that the Spark SQL performance is not very well as we expected.

Software architects
Software developers

?

Example scenario (2)

21,660 issues

466 design decisions

226 structural decisions | 389 behavioral decisions | 166 ban decisions

TUM

Users

SPARK-7393

We want to use Spark SQL in our project, but we found that the Spark SQL performance is not very well as we expected.

Software architects
Software developers

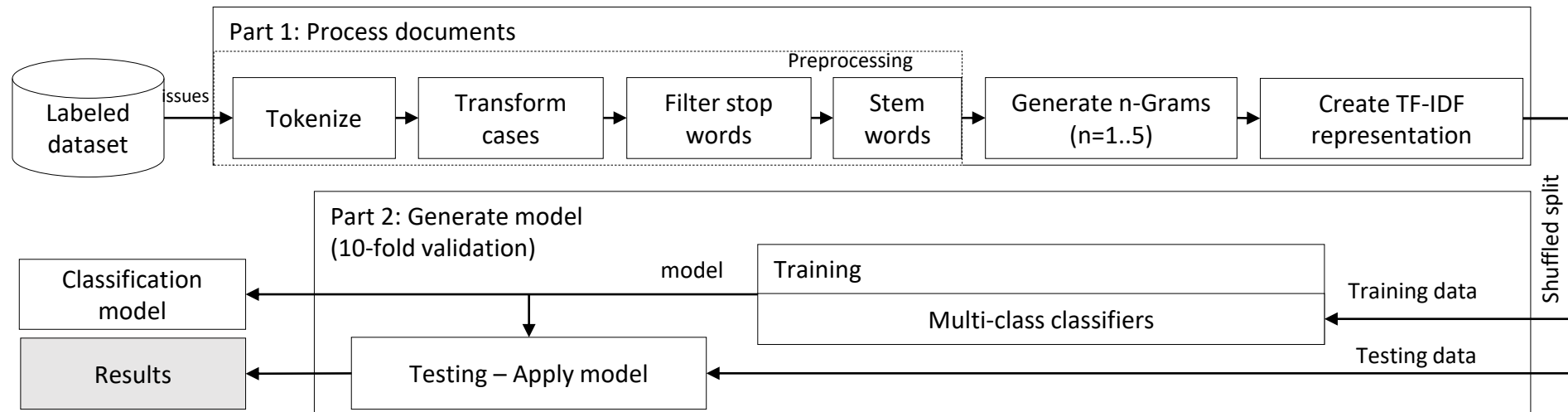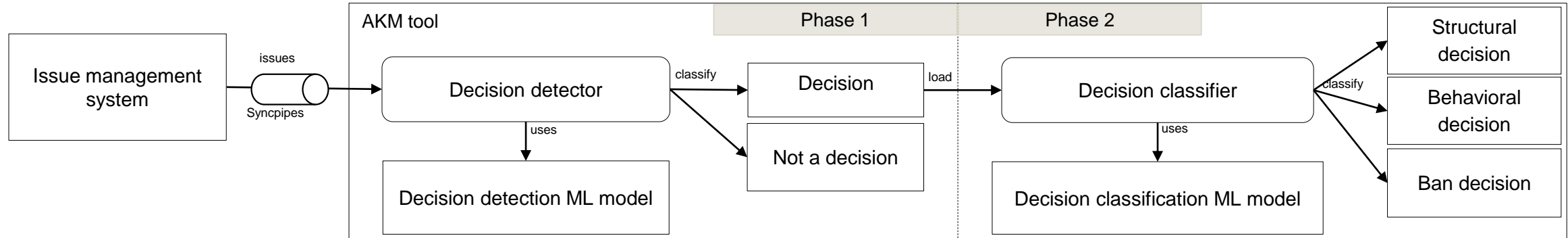"remove unneeded commons httpclient dependencies from pom files in hadoop and sub projects" "In branch and later the patches for various child and related bugs listed in HADOOP most recently including HADOOP HADOOP HADOOP HADOOP and HDFS eliminate all use of commons httpclient from Hadoop and its sub projects except for hadoop tools hadoop openstack see HADOOP However after incorporating these patches commons httpclient is still listed as a dependency in these POM files hadoop project pom xml hadoop yarn project hadoop yarn hadoop yarn registry pom xml We wish to remove these but since commons httpclient is still used in many files in hadoop tools hadoop openstack we ll need to add the dependency to hadoop tools hadoop openstack pom xml We ll add a note to HADOOP to undo this when commons httpclient is removed from hadoop openstack In this was mostly done by HADOOP but the version info formerly inherited from hadoop project pom xml also needs to be added so that is in the branch version of the patch Other projects with undeclared transitive dependencies on commons httpclient previously provided via hadoop common or hadoop client may find this to be an incompatible change Of course that also means such project is exposed to the commons httpclient CVE and needs to be fixed for that reason as well"

reynold xin – 3

sandy ryza – 3

hyukjin kwon – 4

josh rosen – 2

"avoid per record type dispatch in jdbc when writing" "This is similar with https issues apache org jira browse SPARK Currently JdbcUtils savePartition is doing type based dispatch for each row to write appropriate values So appropriate writers can be created first according to the schema and then apply them to each row This approach is similar with CatalystWriteSupport"

djvulee – 2

liwei lin(inactive) – 2

"ugi should contain authentication method" "The UserGroupInformation should contain authentication method in its subject This will be used in HDFS to issue delegation tokens only to kerberos authenticated clients"

marcelo vanzin – 2

Extraction and classification of design decisions from issue management systems

Bhat M.; Shumaiev, K.; Biesdorf, A.; Hohenstein, U.; Matthes, F.: Automatic extraction of design decisions from issue management systems: a machine learning based approach. ECSA 2017.
https://github.com/sebischair/DocClassification

An ontology-based approach for software architecture recommendations

- DBpedia ontology – 4.2 million instances, 685 concepts, 2795 properties

- Automatic extraction and annotation of architectural elements in text
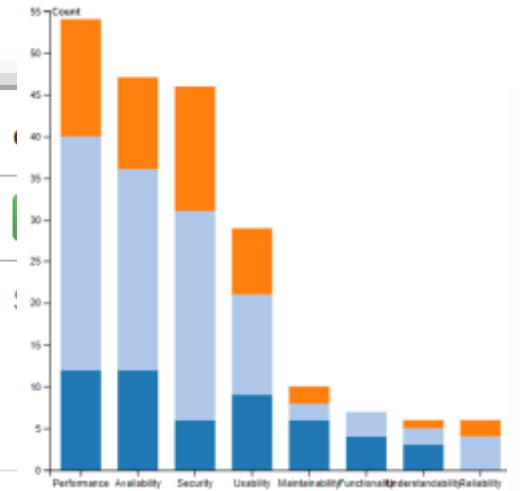- Generate recommendations for alternative options for decision making
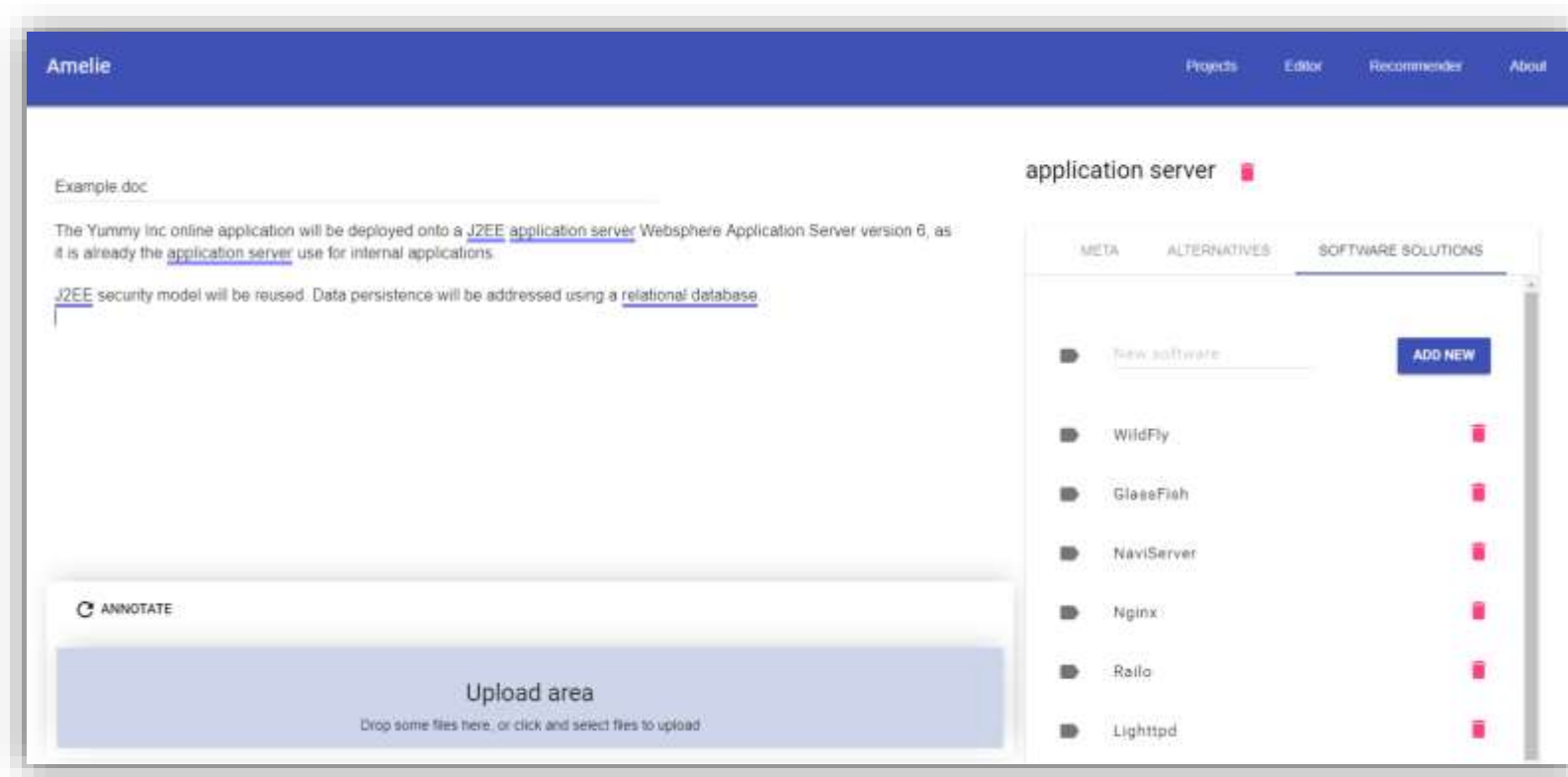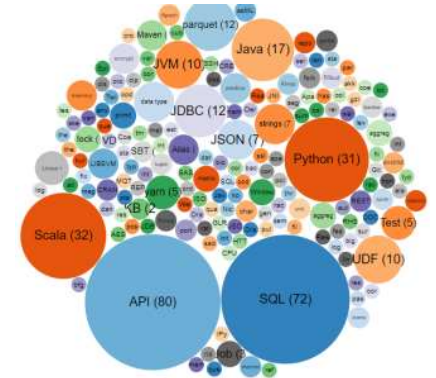
Bhat M.; Shumaiev, K.; Biesdorf, A.; Hohenstein, U.; Hassel, M.; Matthes, F.: An Ontology-based Approach for Software Architecture Recommendations, AMCIS 2017 Boston.
https://github.com/sebischair/akre-server

Why was a design decision made?

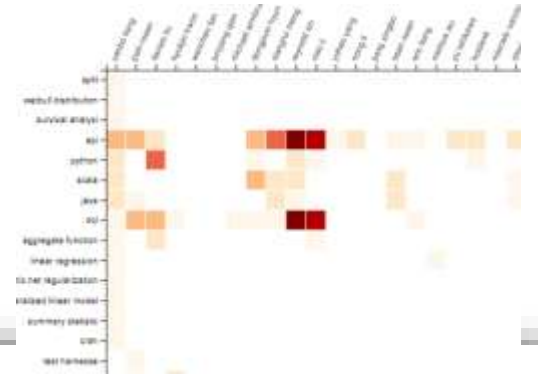|	Task	belongs_to	concepts	Description	design decision	patterns	qualityAttributes	Summary		
☐	HADOOP-13037	Hadoop Common	file system , rest interface , API	The jira proposes an improvement over HADOOP to...	Yes		Security	Refactor Azure Data Lake Store as an independent FileSystem		
☐	HADOOP-12811	Hadoop Common	port , HBase	The HBase s HMaster port number conflicts with Hadoop kms...	Yes		Portability	Change kms server port number which conflicts with HMaster port number	Structural decision	Improvement
☐	HADOOP-13687	Hadoop Common	file system	Currently downstream projects that want to integrate...	Yes		Maintainability	Provide a unified dependency artifact that transitively includes the cloud storage modules shipped with Hadoop	Structural decision	Improvement
☐	HADOOP-12321	Hadoop Common	JVM , thread safe	The new JVM pause monitor has been written with its own...	Yes		Security	Make JvmPauseMonitor an AbstractService	Structural decision	New Feature

DEMO: goo.gl/mgE5mo

Who should be involved in making a design decision?

- How to quantify the architectural expertise of an architect/developer?
- What should be the expertise of an architect/developer to address a design decision?

"remove unneeded commons httpclient dependencies from pom files in hadoop and sub projects" "In branch and later the patches for various child and related bugs listed in HADOOP most recently including HADOOP HADOOP HADOOP HADOOP and HDFS eliminate all use of commons httpclient from Hadoop and its sub projects except for hadoop tools hadoop openstack see HADOOP However after incorporating these patches commons httpclient is still listed as a dependency in these POM files hadoop project pom xml hadoop yarn project hadoop yarn hadoop yarn registry pom xml We wish to remove these but since commons httpclient is still used in many files in hadoop tools hadoop openstack we ll need to add the dependency to hadoop tools hadoop openstack pom xml We ll add a note to HADOOP to undo this when commons httpclient is removed from hadoop openstack In this was mostly done by HADOOP but the version info formerly inherited from hadoop project pom xml also needs to be added so that is in the branch version of the patch Other projects with undeclared transitive dependencies on commons httpclient previously provided via hadoop common or hadoop client may find this to be an incompatible change Of course that also means such project is exposed to the commons httpclient CVE and needs to be fixed for that reason as well"

reynold xin -- 3

sandy ryza -- 3

hyukjin kwon -- 4

josh rosen -- 2

"avoid per record type dispatch in jdbc when writing" "This is similar with https issues apache org jira browse SPARK Currently JdbcUtils savePartition is doing type based dispatch for each row to write appropriate values So appropriate writers can be created first according to the schema and then apply them to each row This approach is similar with CatalystWriteSupport"

djvulee -- 2

liwei lin(inactive) -- 2

"ugi should contain authentication method" "The UserGroupInformation should contain authentication method in its subject This will be used in HDFS to issue delegation tokens only to kerberos authenticated clients"

marcelo vanzin -- 2

DEMO: goo.gl/mgE5mo

Big picture

TLM

| 1 Import project data | 2 Extract design decisions | 3 Annotate architectural elements | 4 Annotate quality attributes | 5 Generate recommendations |

ML model for decision detection and classification

DBpedia

ISO/IEE 25010

Expertise profile

software project

SyncPipes

Software architects
Software developers

Decision classifier — uses

AE annotator — uses

Architectural solutions recommender — uses

Rationale extractor — uses

Expert recommender — uses

project data

documents — design decisions

documents — annotated document

architectural element (AE) — recommendations

design decisions — quality attributes

design decisions — architects & developers

Meta-model based AKM system

DEMO
goo.gl/mgE5mo

Bhat M.; Shumaiev, K.; Matthes, F.: Towards a framework for managing architectural design decisions. ECSA 2017.

M.Sc.
Manoj Mahabaleshwar

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel   +49.89.289.18162
Fax   +49.89.289.17136

manoj.mahabaleshwar@tum.de
wwwmatthes.in.tum.de
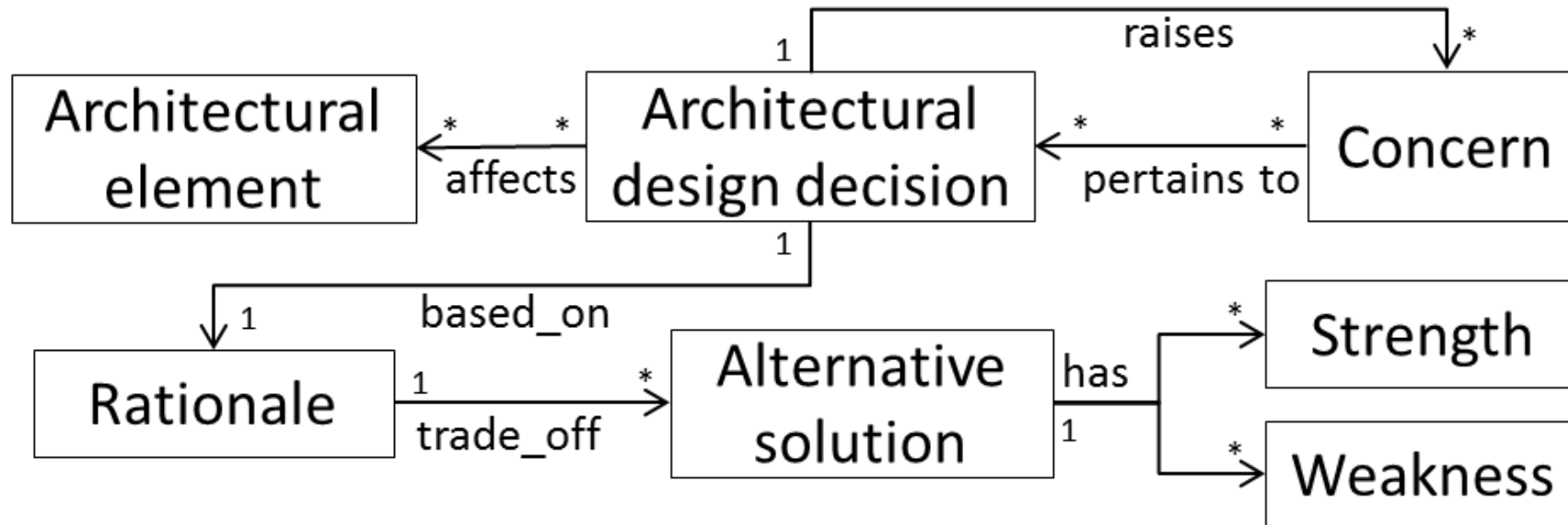
Research questions

RQ 1: How to consolidate project data from disparate information systems?

RQ 2: How to extract design decisions from textual information?

RQ 3: How to identify architectural elements and recommend alternative choices to support decision making?

RQ 4: How to identify the rationale behind a design decision?

RQ 5: Who should be involved in making a design decision?

SyncPipes

A platform for implementing reusable adapters

JIRA
GitHub Issues
Microsoft Project
Enterprise Architect

SyncPipes Configurator

Input configuration | Sync configuration | Output configuration

SyncPipes Mapper

concept mapper
concept mapper
attribute mapper

Get Types and Properties
Get Types and Properties

SyncPipes Execution

SyncPipes Synchronizer

Tool A
Data model A

SC model
SC KB

establish connection
establish connection
Get Entities
Create Entities
Update Entities

https://github.com/sebischair/syncpipes-server,
https://github.com/sebischair/syncpipes-client

Bhat, et al.: Meta-model Based Framework for Architectural Knowledge Management, SAGRA workshop at ECSA 2016

SocioCortex for architectural knowledge management

Source tools

Non-structural data

Client Apps

SyncPipes | IR-application | Model designer | Rules manager | Value Added Services

Platform

REST API

Access control

Rule engine (Reasoning knowledge)

Events & notifications | Model-based expressions

Meta-model

Static knowledge model | Dynamic knowledge model

Static knowledge base

Dynamic knowledge base
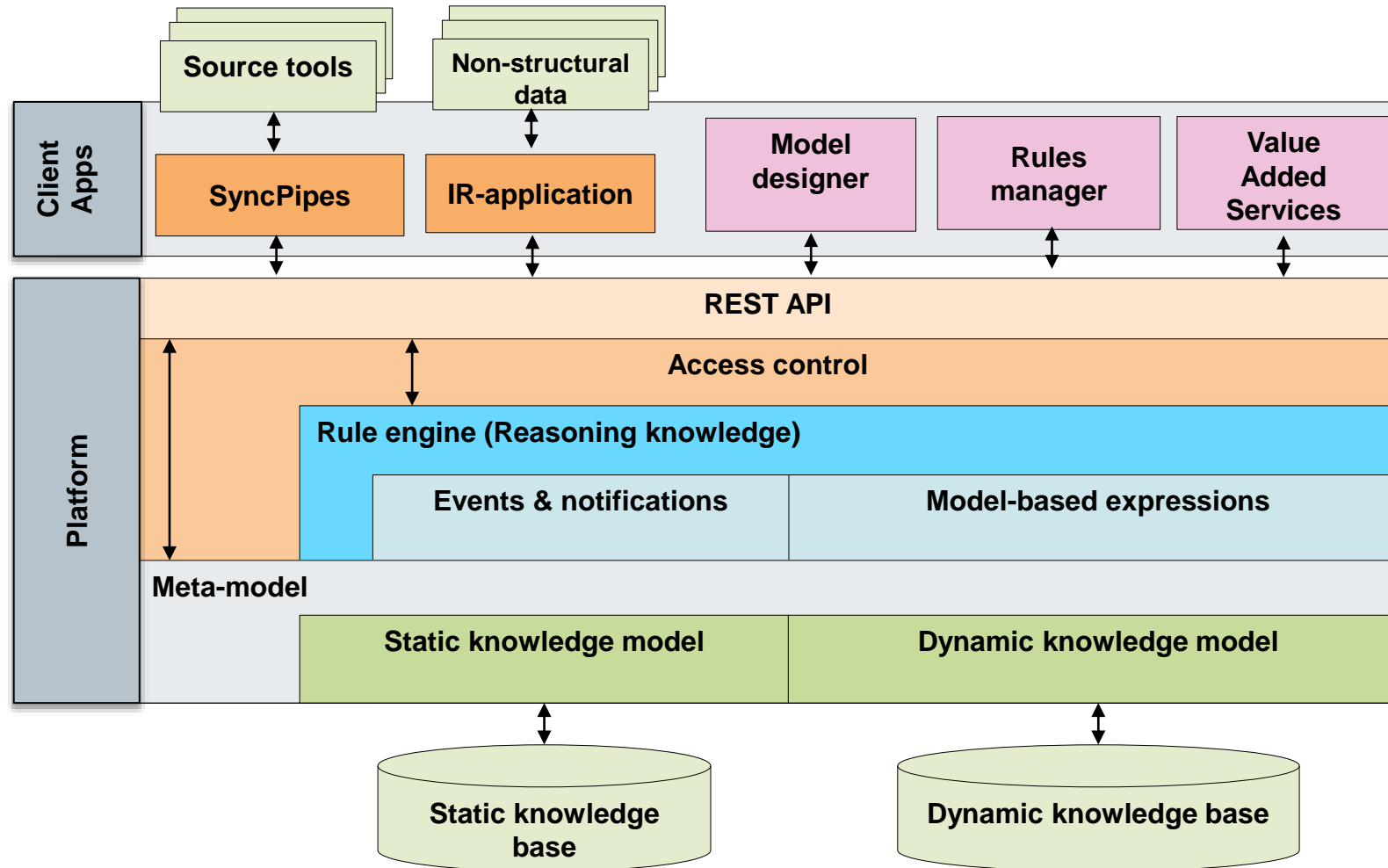
Bhat, M.; Shumaiev, K.; Biesdorf, A.; Hohenstein, U.; Hassel, M.; Matthes, F.: Meta-model Based Framework for Architectural Knowledge Management, SAGRA workshop at European Conference on Software Architecture 2016, Nov. 28 - Dec. 02, 2016, Copenhagen, Denmark
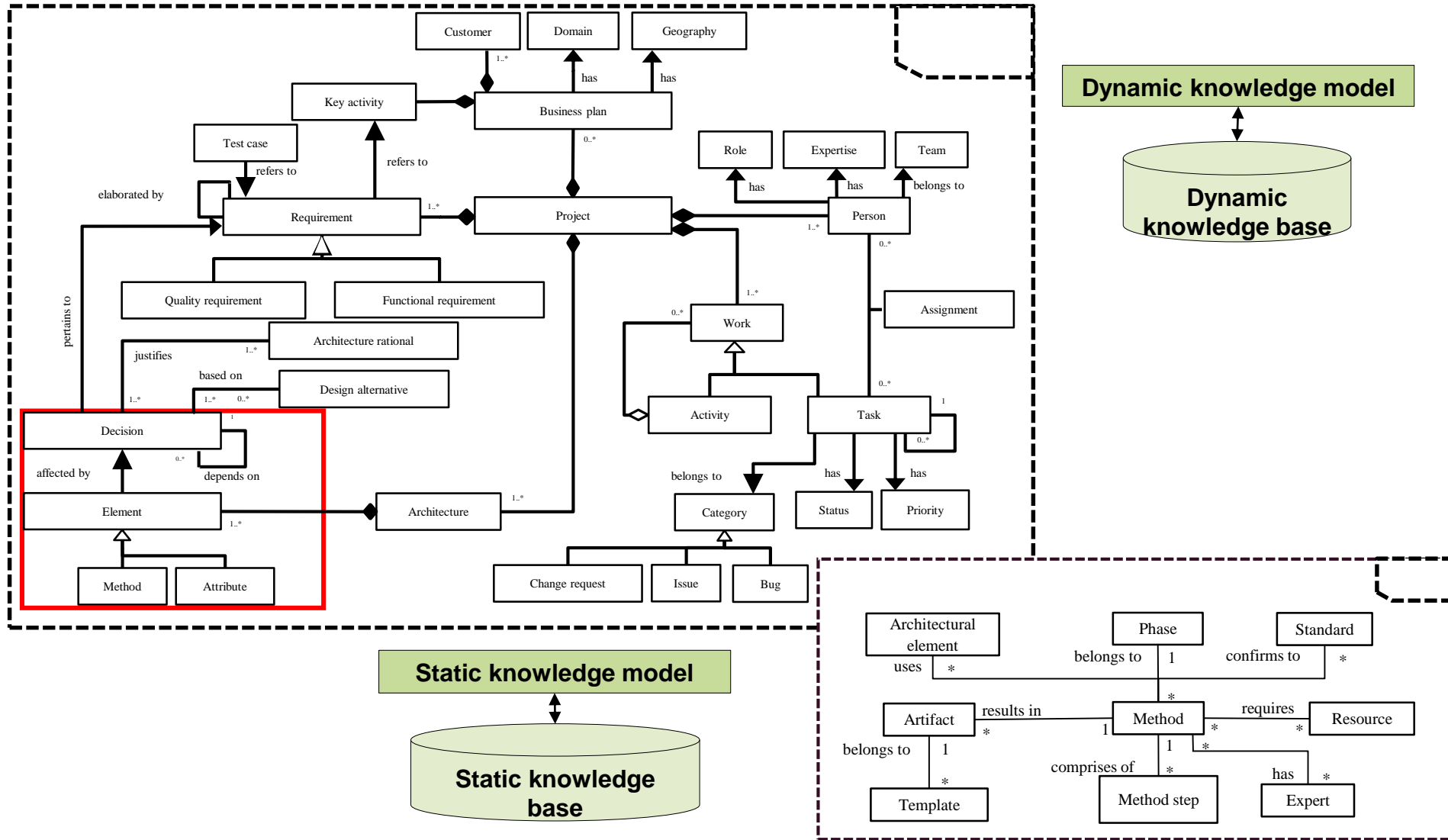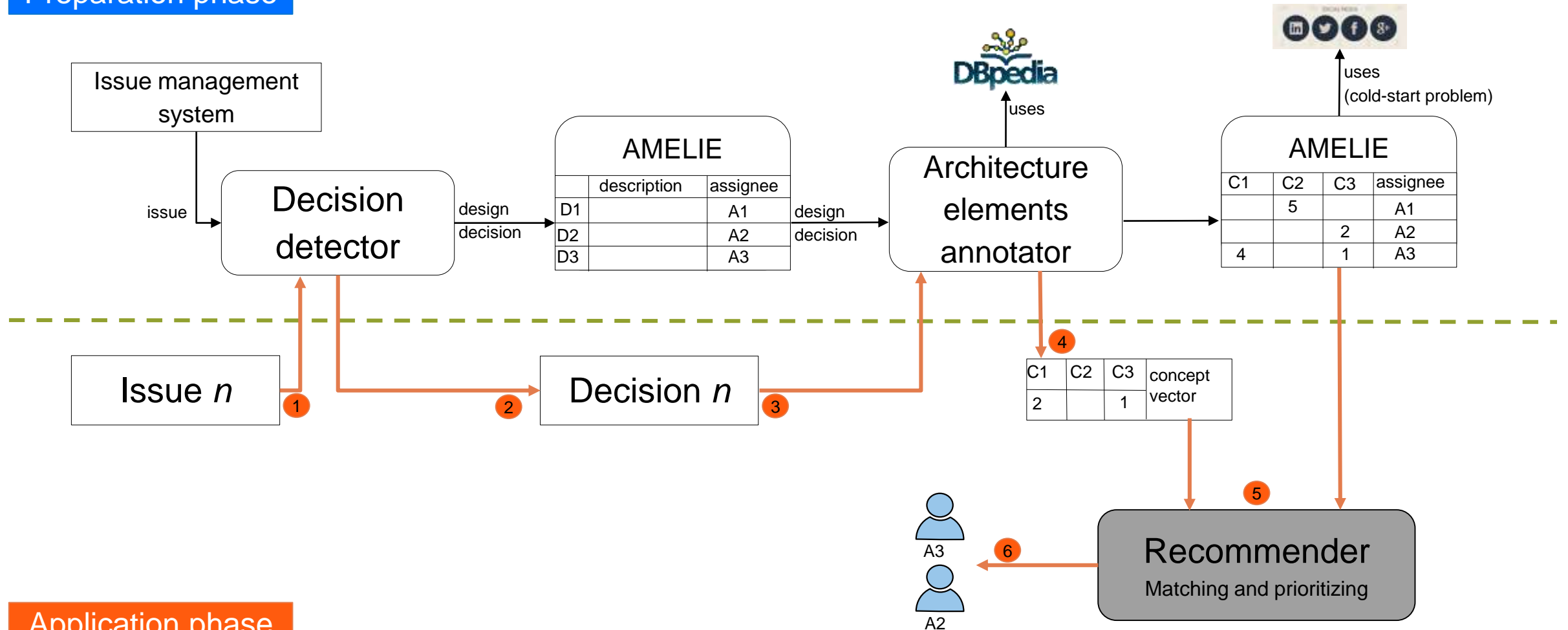
Who should be involved in making a design decision?

RQ 5

Preparation phase

Issue management system

AMELIE

	description	assignee
D1		A1
D2		A2
D3		A3

DBpedia

uses

Architecture elements annotator

uses
(cold-start problem)

AMELIE

C1	C2	C3	assignee
	5		A1
		2	A2
4		1	A3

issue → **Decision detector** → design decision → design decision →

Issue *n*

①

②

Decision *n*

③

④

C1	C2	C3	concept vector
2		1	

⑤

A3

⑥

A2

Recommender
Matching and prioritizing

Application phase

Which similar design decisions were made in the past?

Given a new design decision, search the knowledge base for *similar* earlier made design decisions.

Research questions:

- How to identify similar design decisions?
- What are the context parameters that needs to be considered?

Analyze the alternatives for performing text similarity

- Syntactic similarity
- Machine learning algorithms for unsupervised clustering
- Semantic similarity –NLP based (using tag trees)

https://wwwmatthes.in.tum.de/pages/l2inbfu3sbe7/Master-Thesis-Prateek-Bagrecha

Heuristics in decision making

Hypothesis:

Software architects do not always take rationalistic approach (best-fit) during decision making (in general humans are not "maximizers") but favor naturalistic decision making (*heuristic-based, first satisfiable solution, first-fit*)

Research questions:

What are the heuristics used by software architects and developers during the decision-making process?

Examples:

- Past experience - I have successfully used X in the past, so I will use it in this project
- Project constraints - My team has experience in Java, so we will use Java for implementation
- Project's domain specific - In-memory databases are efficient for managing user sessions in web applications
- Trending technology – OAuth, material design, microservices…

https://wwwmatthes.in.tum.de/pages/19gmopufr04wi/Master-Thesis-Akash-Manjunath

Who did what, why and when?

Existing work on ADD models **do not** consider **user preferences, project context,** and **heuristics** to support the decision-making process

AIM

- Extract *who* took the decision, how long did it take to implement, complexity of tasks
- Extract architects' *preferences* related to technologies and types of issues handled to build user profiles.

RESULT

- Combining user profiles with the ADD model should support the recommendation of
  - *Who* should be responsible to address a specific concern?
  - What is the *complexity* of addressing a specific concern?
  - What are the *alternatives* available to address a problem?
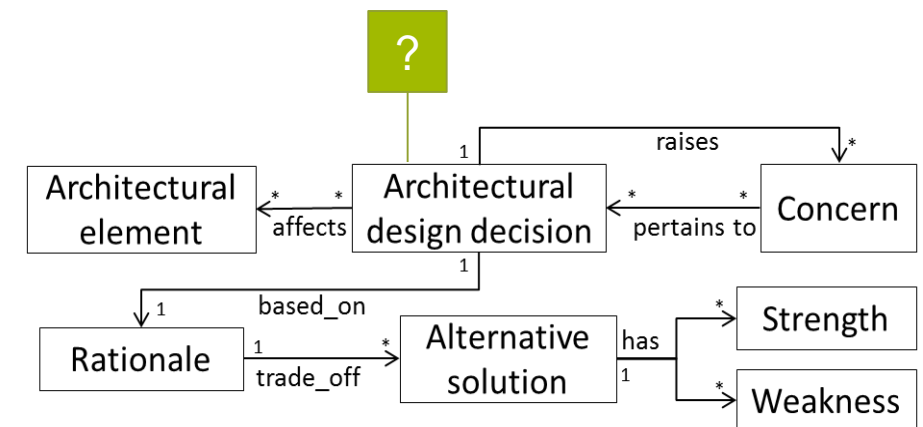
How are design decisions interrelated?

- Focus on extracting relationships between design decisions (influences, refinedBy, decomposesInto, compatibleWith, incompatibleWith)
  - Important for
    - Documentation - specifying constraints on design decisions
    - Communication - visual representation of the constraint graph model
    - Inferring the complexity for addressing similar design decisions

Analyze the applicability of model-based expression language (MxL) in SocioCortex

Decision topic

contains

Design decision

influences
refinedBy
decomposesInto
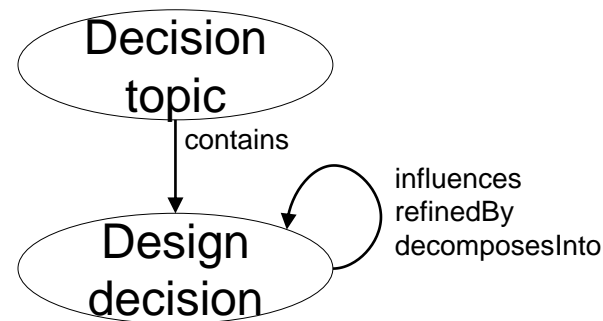
Reference model:
Zimmermann, Olaf, et al. "Managing architectural decision models with dependency relations, integrity constraints, and production rules." *Journal of Systems and Software* 82.8 (2009): 1249-1267.

Timeline

9.17	10.17	11.17	12.17	1.18	2.18	3.18	4.18	5.18	6.18	7.18	8.18	9.18	10.18	11.18	12.18

Similar DD (Master thesis)

Heuristics in decision making? (master thesis)

Who? (wip. focus on publication)

Vacation

Integration + Prepare case studies

Evaluation

Thesis writing + review