Creating Software Architecture Documentation for MediaWiki Software

Master‘s Thesis Final Presentation

14.09.2015, Uliana Bakhtina

Software Engineering für betriebliche Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München

wwwmatthes.in.tum.de

Agenda

1. Motivation
2. Research Questions
3. Methodology
4. Implementation
5. Outlook


Motivation

MediaWiki

- MediaWiki is a software that powers Wikipedia
- MediaWiki was introduced in 2002 and has been constantly developed further by an active volunteer community
- The architecture of MediaWiki has been often determined by initiatives and requests from the community and evolved significantly over time
- The documentation of this architecture was however performed very scarcely

Closing this gap would provide the community with a resource to educate its new members and foster the further contribution from volunteer developers


Research Questions

Research Questions

- **RQ1:** Who are the **stakeholders** for software architecture documentation of MediaWiki?
- **RQ2:** What are the **problems** in available documentation?
- **RQ3:** What are the **steps** to be taken to **understand** the existing software?
- **RQ4:** What are the **steps** to be taken to **produce the documentation.** What kind of **models** and **tools** to use?
- **RQ5:** After the documentation is performed: Did the produced documentation meet the **requirements** and could it bring added **value** to the stakeholders?


Methodology

Identified Work Packages

Literature research

Requirements' elicitation

Analysis of MediaWiki software

Documentation of MediaWiki software

Evaluation

Methodology

Requirements' elicitaion	Analysis of MediaWiki software	Documentation of MediaWiki software	Evaluation
• Identify the stakeholders • Derive requirements from stakeholders • Derive requirements from identified problems in current documentation	• Analyze the software from user's perspective • Analyze use case scenarios • Study development documentation and interview stakeholders	• Select areas for documentation • Define structure for documentation • Select notation • Select tools	• Evaluate the documentation based on the achievement of initial requirements • Interview the stakeholders regarding the added value of produced documentation


Implementation

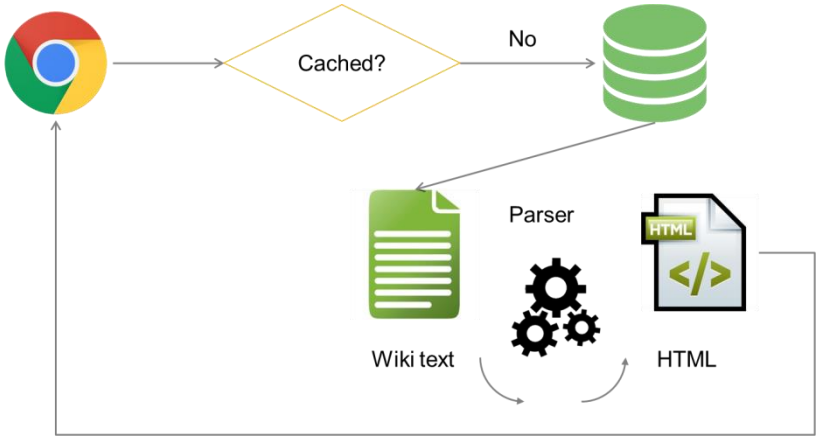
Requirements' elicitation

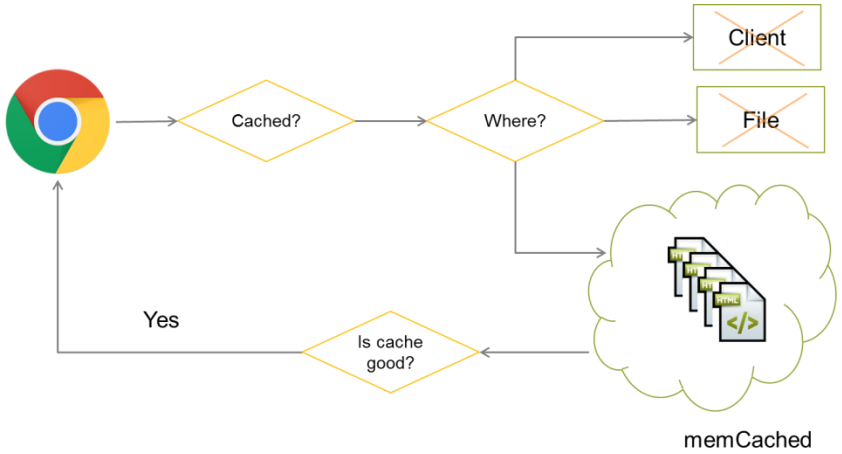
- **Identified stakeholders**
  - New software developer - a person who is not familiar with MediaWiki software, but intends to work on its core
  - Software developer - a developer of MediaWiki software
- **Stakeholders' requirements**
  - Wiki format for the documentation
  - Editable diagrams
- **Requirements derived from the identified problems in the current documentation**
  - Defined approach for the description of system's architecture
  - Overview of the interaction between identified architectural parts
  - Common structure for the description of architectural parts
  - Sufficient implementation details for a new developer
  - Entry point for the user of software architecture documentation with overview of the system and available documentation
  - Structured navigation through the documentation
  - Graphical visualizations to support the documentation

Analysis of MediaWiki Architecture

- **Initial activities**
  - Analysis of the functionality of MediaWiki from user's perspective
  - Local installation of MediaWiki
  - Study of available documentation and interview with developers

- **Scenario-based analysis**
  1. Request for a non-cached page
  2. Request for a cached page
  3. Request for a special page

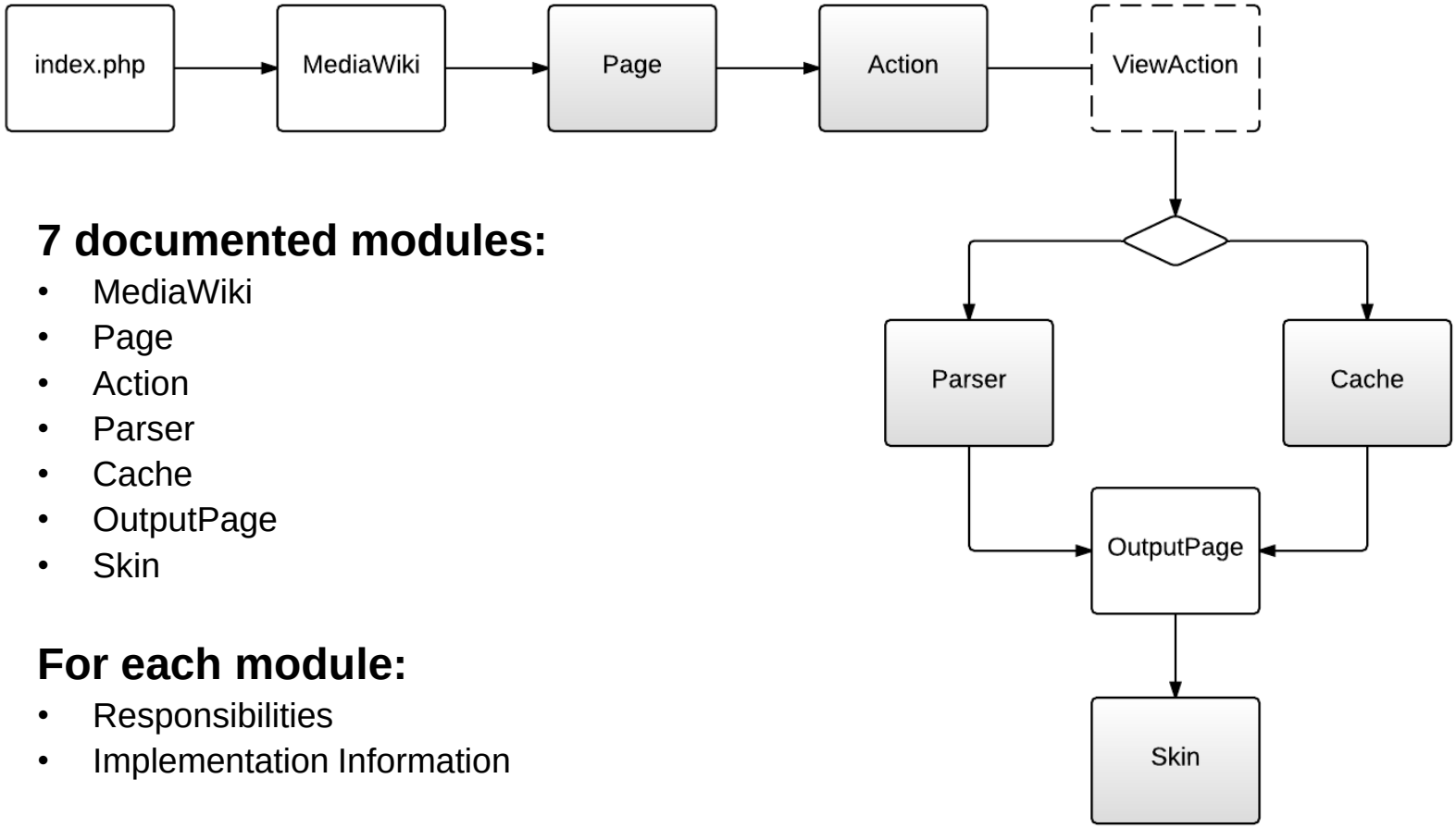




Documentation of MediaWiki Architecture

- **Areas for documentation**
  - Static structure – Description of Modules
    Module is an implementation unit of software that provides a coherent set of responsibilities.
  - Dynamic behavior – Sequence Diagrams
- **Structure of module description**
  - Name – the name of module
  - Responsibilities - role of the module in the system
  - Implementation information - details related to the module's implementation that is relevant for managing its development
- **Notation**
  - UML class diagrams
  - UML sequence diagrams
  - Flow charts
- **Tools**
  - Lucidchart
  - MediaWiki

Architectural Modules of MediaWiki



7 documented modules:

- MediaWiki
- Page
- Action
- Parser
- Cache
- OutputPage
- Skin

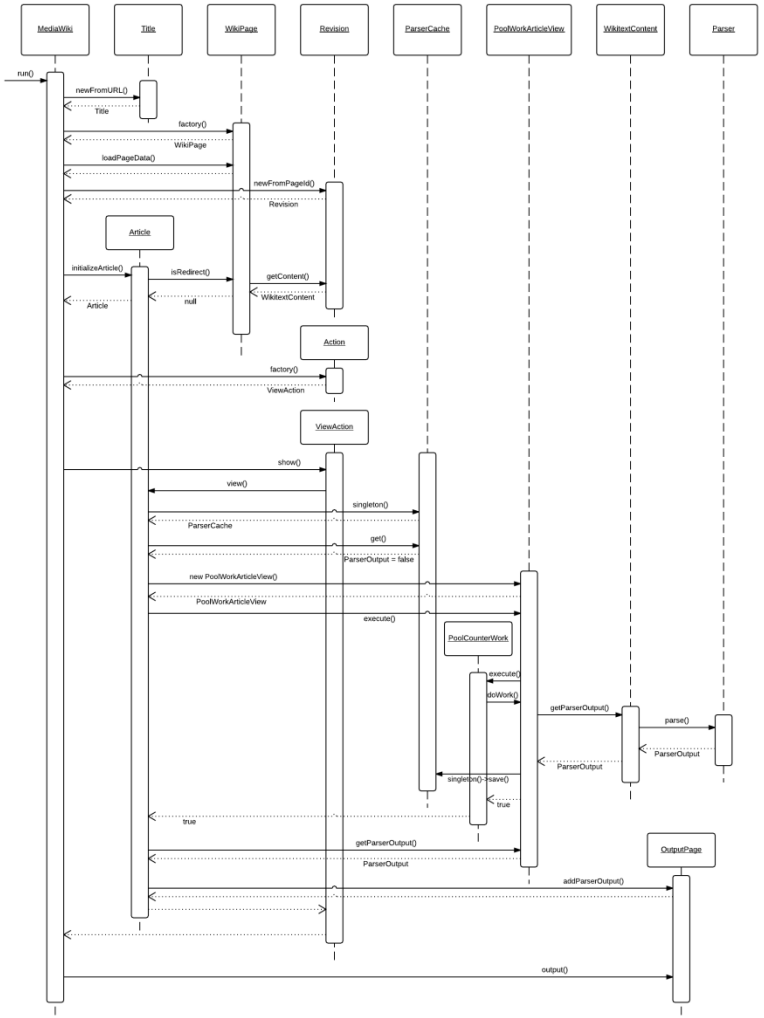
For each module:

- Responsibilities
- Implementation Information

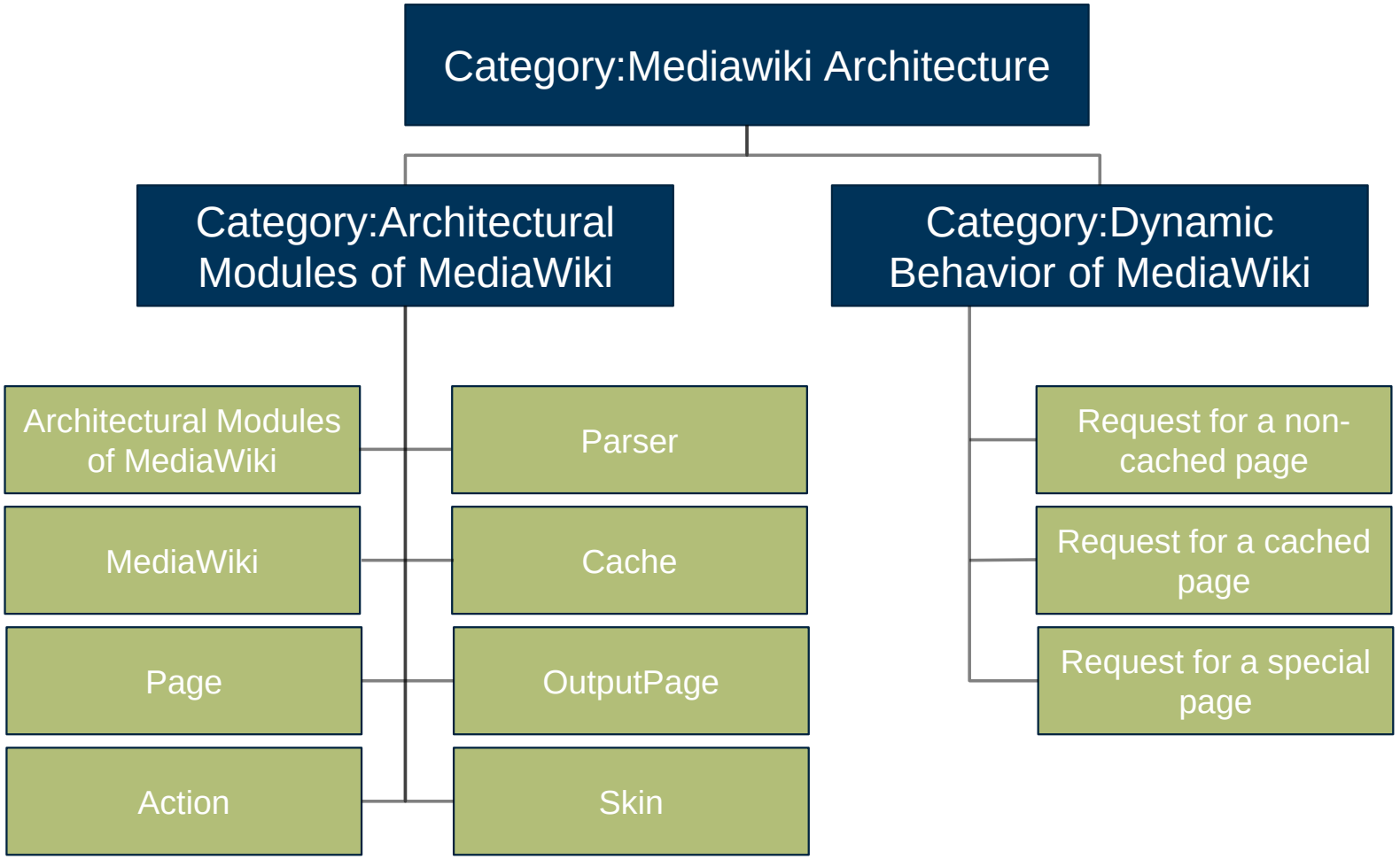
Dynamic Behavior of MediaWiki

3 UML sequence diagrams and description of the requests

- Request for a non-cached page
- Request for a cached page
- Request for a special page



Resulted Wiki



Evaluation

All 9 requirements have been fulfilled

Requirements	Fulfilled
Wiki format for the documentation	✓
Editable diagrams	✓
Defined approach for the description of system's architecture	✓
Overview of the interaction between identified architectural parts	✓
Common structure for the description of architectural parts	✓
Sufficient implementation details for a new developer	✓
Entry point for the user of software architecture documentation with overview of the system and available documentation	✓
Structured navigation through the documentation	✓
Graphical visualizations to support the documentation	✓

Evaluation

- **Interview with stakeholders**
  - 11 questions
  - Emails to the stakeholders
  - 2 answers

- **Summary**
  - The documentation is good and more detailed than what was available before. Level of detail is right
  - The documented modules are relevant and useful to new core developers and this kind of documentation was absent before
  - Selected methodology to use module views for static structures is a good choice
  - Description of dynamic behavior is very useful

- **Suggestions**
  - Explanation of intended architecture and divergence from reality
  - SVG format for diagrams
  - Communication diagrams instead of sequence diagrams
  - Further feedback from a larger audience like wikitech-l mailing list
  - Placement of documentation on www.mediawiki.org
  - Primary copy of the documentation in the docs/ directory of the git repository of MediaWiki


Outlook

Outlook

- **Document further modules to cover the entire system**
  - Database
  - Special Page
  - API
  - Media
  - Internationalization
  - …
- **Document further cases describing dynamic behavior**
  - Request for uploading and image and creating thumbnail for it
  - API requests
  - Other Special Page requests
  - …
- **Document architectural decisions and divergence of actual architecture from desired architecture.**
- **Improve maintainability of produced documentation by placing it in text files format under docs folder in MediaWiki source code**

Thank you for your attention. Questions?

Uliana Bakhtina
B.Sc.

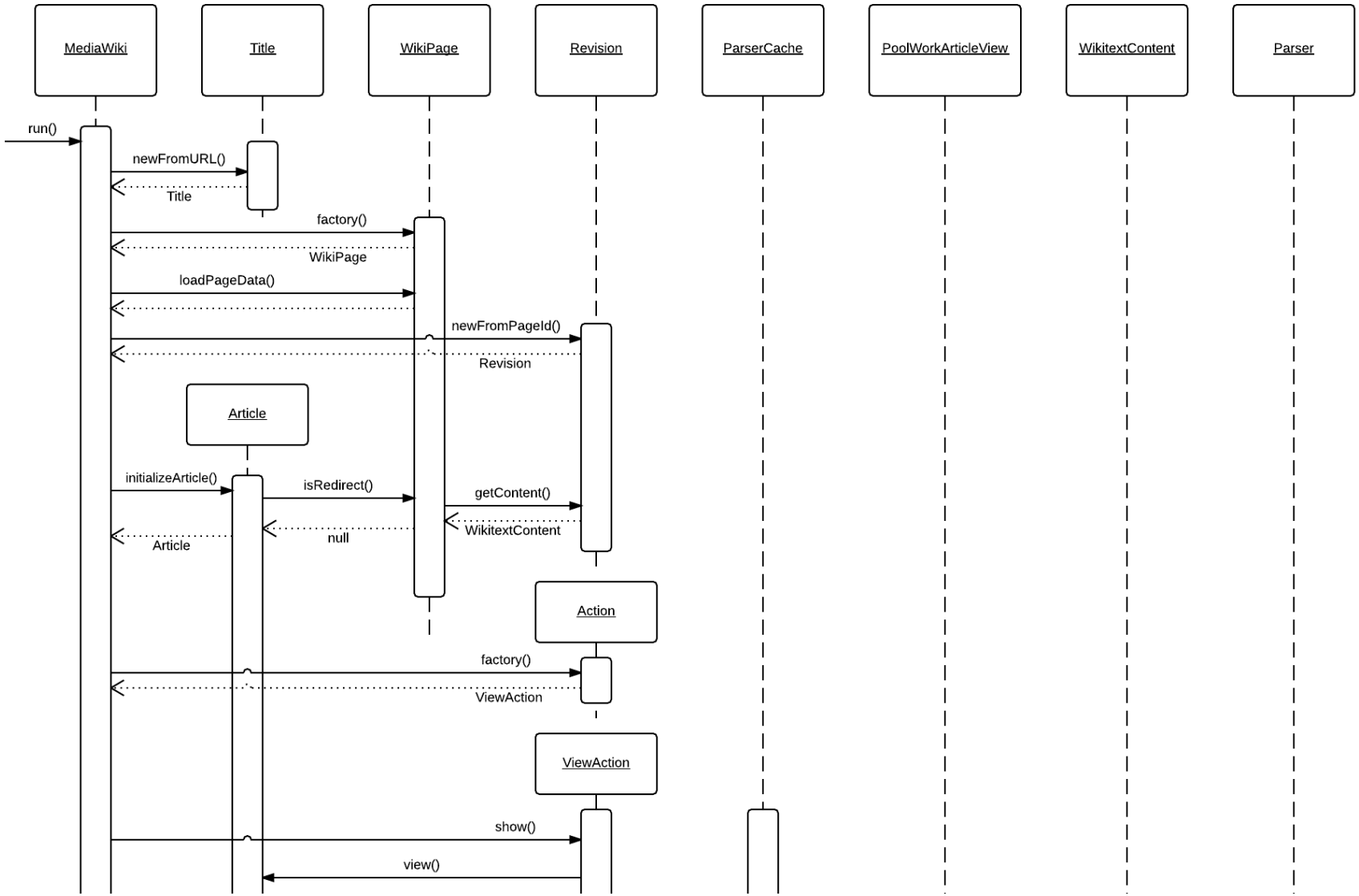
Technische Universität München
Department of Informatics
Chair of Software Engineering for
Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

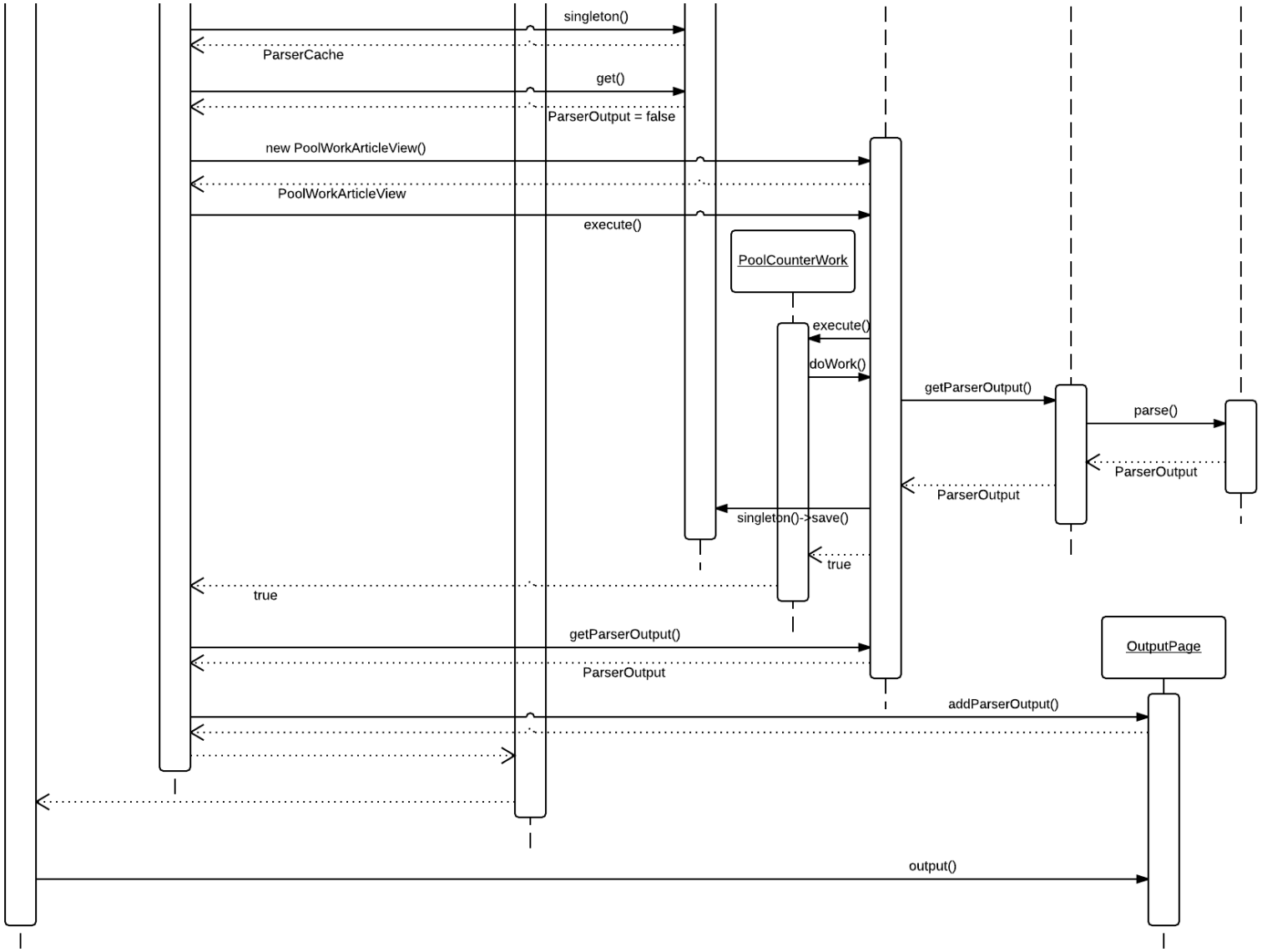
Tel +49.89.289.17132
Fax +49.89.289.17136

uliana.bakhtina@tum.de
wwwmatthes.in.tum.de

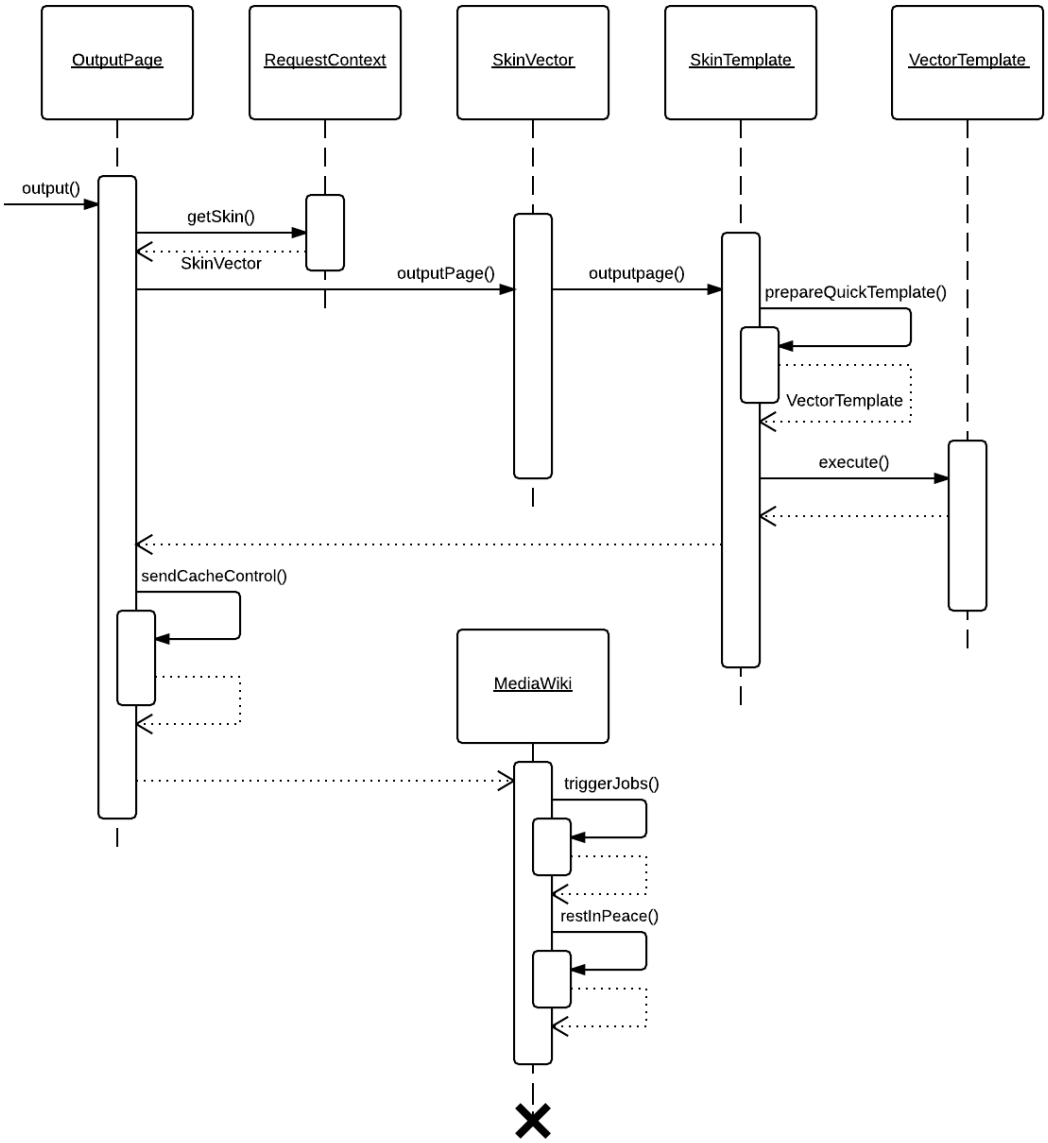
Request for a non-cached page (1)



Request for a non-cached page (2)



Request for a non-cached page (3)



Scenario details example

Request for a non-cached page	
Scenario Description	A request is sent to view a page that has not been cached before and thus an HTML output needs to be generated by the application
Objectives	• Analyze the work of parser for converting wikitext into proper HTML • Analyze the assembling of the page using MediaWiki's skins mechanism • Analyze the execution flow and involved classes
Special Setup	The requested page has a lot of different wikitext formatting in order to see the parsing of these elements into HTML
Architectural Modules	Parser Skin
Produced Documentation Artifacts	• Documentation of Parser module • Documentation of Skin module • Sequence diagram for the request for a non-cached page and explanation for it