AISOP

AI-Supported Observation of E-Portfolios

https://aisop.de

Conception of a System Architecture for the AI-Supported Observation of e-Portfolios

Authors

Wolfgang Müller

Sandra Rebholz

Paul Libbrecht

Alexander Gantikow

Andreas Isking

Thierry Declerck

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Baden-Württemberg

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

Pädagogische Hochschule Weingarten

Introduction

E-portfolios are digital works that learners create to express their knowledge and the process of acquiring this knowledge. As a composition, an e-portfolio lets learners express themselves with digital artefacts (texts, graphics, videos, ...) using a vocabulary that is both accessible to them and to their readers. As a digital composition, e-portfolios enjoy the ease of re-use to create a representation of their learning process and knowledge that is their own. The AISOP (AI-supported observation of e-Portfolios) project aims at supporting students' learning by the means of e-portfolios and by applying AI-based tools and methods to analyse their e-portfolios automatically. Based on the analysis results and their representation in learning dashboards, both the students get individual support in the e-portfolio creation process and the teachers get support in assessing the created artefacts.

In this report, we present a system architecture and technical infrastructure that allows for a seamless integration of a standard e-portfolio platform, suitable AI-based tool chains, and interactive dashboard applications for students and teachers. This is followed by a description of how the architecture's components interact in a typical analysis workflow. Furthermore, we examine the development of an integrated knowledge architecture for guiding the semantic analysis of e-portfolios considering two different knowledge resources. The proposed architecture is the first draft of an AI-supported e-portfolio analysis system to be used in real-life scenarios at the University of Education in Weingarten.

An Infrastructure to Process e-portfolios

E-portfolios considered in this project are written by students mostly using the university-internal portfolio platform Mahara. The open-source project is widely used to offer editing and sharing possibilities that are typically necessary to create portfolios in the form of a series of HTML pages that can be shared with others. Once the students have created their individual portfolios, we propose that the infrastructure employs a Mahara plug-in to deliver, after authorization, the portfolio content to the AISOP system as sketched in Figure 1.

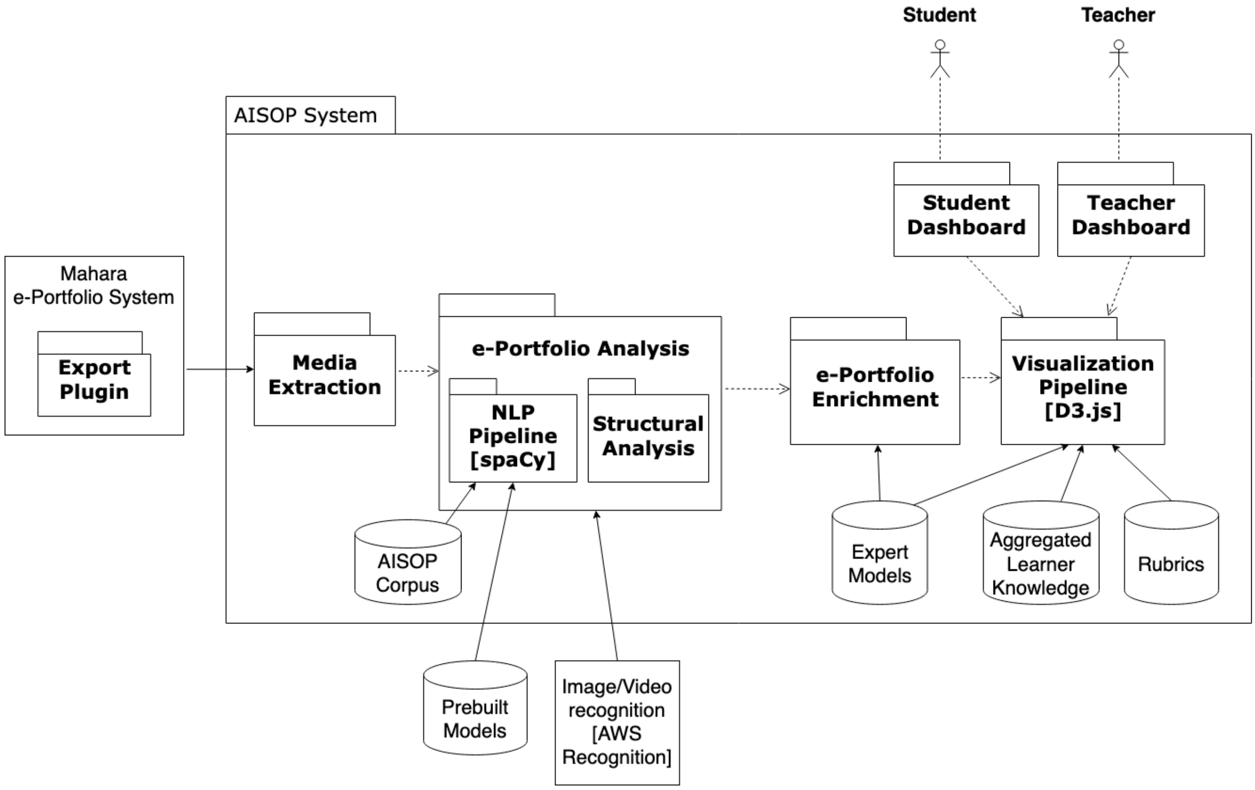

Figure 1: Sketch of the software architecture for the AISOP system.

Based on the premise that a portfolio is a series of HTML files with embedded media (images and videos), it can enter three different processes which are sketched in this picture:

- The mere display: A human reads the portfolios for evaluation, understanding and navigation: The HTML needs to be faithfully presentable and be mostly equivalent to what the students are seeing when editing and viewing (in Mahara): This assumes that the exported content of the portfolio is not a mere collection of texts or a fully laid out PDF, but a set of interactive documents including multimedia artifacts and links to web-based resources.
- The enrichment: By enriching the HTML using the result of text-analysis (e.g. entities recognized, topics identified, annotations on each paragraph) and the result of media analysis (yielding enriched caption texts using the character and voice recognition functions of stock recognition web-services), dashboards can provide a visual synthesis of the content. They do so either of single portfolios or of groups of portfolios (e.g. for comparison or for analysis of the results of a cohort, for example for course quality analysis).
- The inclusion: Portfolios can be copied into the corpus of portfolios for annotations. After a first anonymization, an annotation process is operated (e.g. using the prodigy*2). The annotated corpus allows creating pipelines that can recognize relevant entities and topics. Provided all authorizations are obtained, and the anonymization is ensured, the corpus can be shared for other institutions to be able to create the same pipelines.

These processes are provided by online services that can analyse e-portfolios of a cohort or of a single user, visualizing the result and offering navigation services. From this sketch, one can see that the infrastructure is made of web-servers and web-server integrations, as well as Natural

Language Processing (NLP) pipelines and their creation processes. Some components are quite specific to the environment (such as the extraction of portfolios from Mahara) but some components are likely reusable in different contexts (e.g., the enrichment of HTML by the text-analysis).

In order for trust to be maintained and the privacy levels required in institutions such as our university of education, a detailed analysis of the sharing authorizations is required. Students should only use the AISOP services if they trust the processes of this architecture, and the right to be forgotten should be guaranteed. One of the milestones of trust that we shall be able to reach is when the maintainers of the Mahara system of the university will consider the authorization system and our architecture as having sufficient measures to be deployed in their system, a productive system used by hundreds of persons every year.

The User's Perspective: Portfolio Analysis with AISOP

In order to illustrate how the components of the proposed architecture interact in a typical analysis workflow, the following scenario describes how a user carries out an AI-based analysis of a portfolio:

The user logs in into the university's portfolio platform (Mahara). At the University of Education in Weingarten, users first log in into the Learning Content Management System 'moopaed' to access Mahara (Single Sign On). Subsequently, they access the Mahara group workspace, which was created specifically for their course team. A conventional hyperlink will redirect them to the external AISOP application. In order to access the AISOP system, users must follow a multi-step authentication and authorization process based on the OAuth 1.0a protocol: as a first step, the users are asked to authorize AISOP to access their user data as well as the portfolios they have access to. In case of a positive confirmation, they are redirected to AISOP where they are now logged in. The system displays a list of portfolios to the user. They can now select a portfolio to deliberately trigger its analysis. As a result, AISOP requests the export of the portfolio from Mahara, downloads it, and sends it to the AI-based analysis pipeline. The user is informed about the process and can finally inspect the results of the analysis in a dashboard. Once a student has made changes to a portfolio, users can initiate the analysis process again. This process is depicted in figure 2.

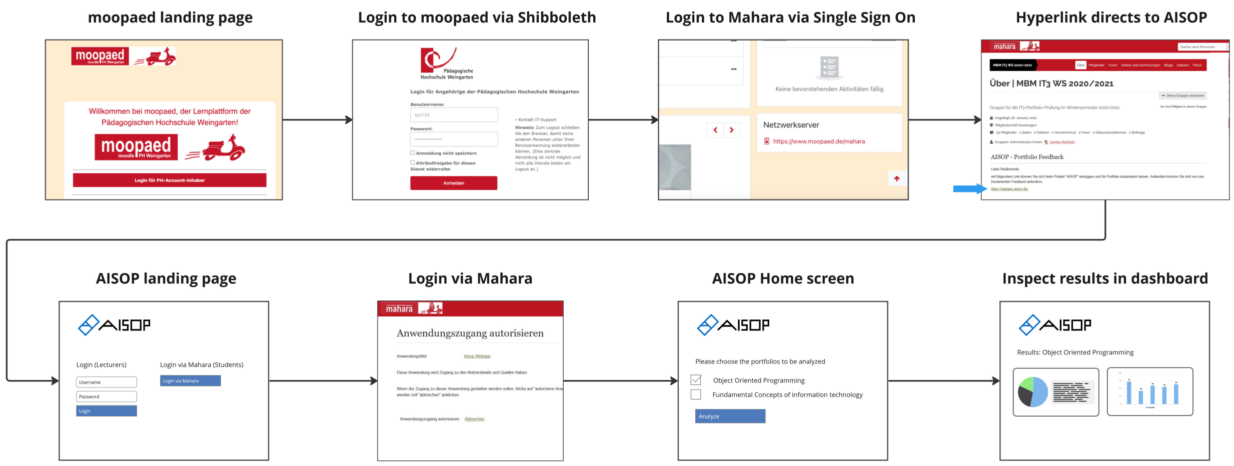

Figure 2: Sequence of screens that the users of the system will see in order to authorize access to portfolios and obtain their analysis.

Towards a Knowledge Architecture for AISOP

In the context of AISOP, dealing with the automated analysis of e-portfolios [1], we are investigating the development of an “onto-terminology” for guiding the semantic analysis of such e-portfolios. The term “onto-terminology” is borrowed from [2], as the here presented approach pursues very similar goals: establishing a (multilingual) domain terminology within an ontological framework.

Two different knowledge resources are considered in this context and have to be unified. The first one, external to AISOP, is the “Computer Science Ontology (CSO)” (see https://cso.kmi.open.ac.uk/home), as the courses in which the e-portfolios are produced are dealing with computer science and programming.

CSO was automatically generated by an information extraction system that is applied to relevant scientific literature, and manually curated. While this knowledge resource is very relevant, we note that it is rather a flat ontology (very close to a taxonomy in fact), and it includes only 8 properties, one of those being foreseen for human readability of the classes or property names (the property rdfs:label). We also note that it contains no (domain) terminology. Figure 3 displays an example of the information encoded in CSO:

```
<https://cso.kmi.open.ac.uk/topics/function_point>
   ns0:relatedEquivalent <https://cso.kmi.open.ac.uk/topics/function_points> ;
   ns0:superTopicOf <https://cso.kmi.open.ac.uk/topics/function_point_analysis> ;
   ns1:label "function point"@en ;
   rdf:type ns0:Topic ;
   ns0:preferentialEquivalent <https://cso.kmi.open.ac.uk/topics/function_point> ;
   ns2:sameAs <http://dbpedia.org/resource/Function_point> ,
       <http://www.wikidata.org/entity/Q1277601> ,
       <http://rdf.freebase.com/ns/m.0bwtzg> ,
       <http://yago-knowledge.org/resource/Function_point> ;
   ns3:relatedLink <https://academic.microsoft.com/#/detail/159765158> ,
       <http://en.wikipedia.org/wiki/Function_point> .
```

Figure 3: An example taken from the Computer Science Ontology

The second resource type considered are concept maps representing the knowledge to be conveyed in the courses, which are conceived manually by teachers of the University of Education Weingarten. The concepts are encoded as instances of Cmap (see https://cmap.ihmc.us/) in a graphical way yielding such graphics as in Figure 4.

An example of current content included in such a Cmap instance, in an XML export, is given in Figure 5, where we observe that terminology-relevant content (definitions, lexical semantic relations, information about the language in use, etc. is given. Here we note that this heterogeneous information is not distributed over specialised features, but stored informally under the feature “long-comment”. It would also be desirable to mark explicitly the language used for the definitions, etc.

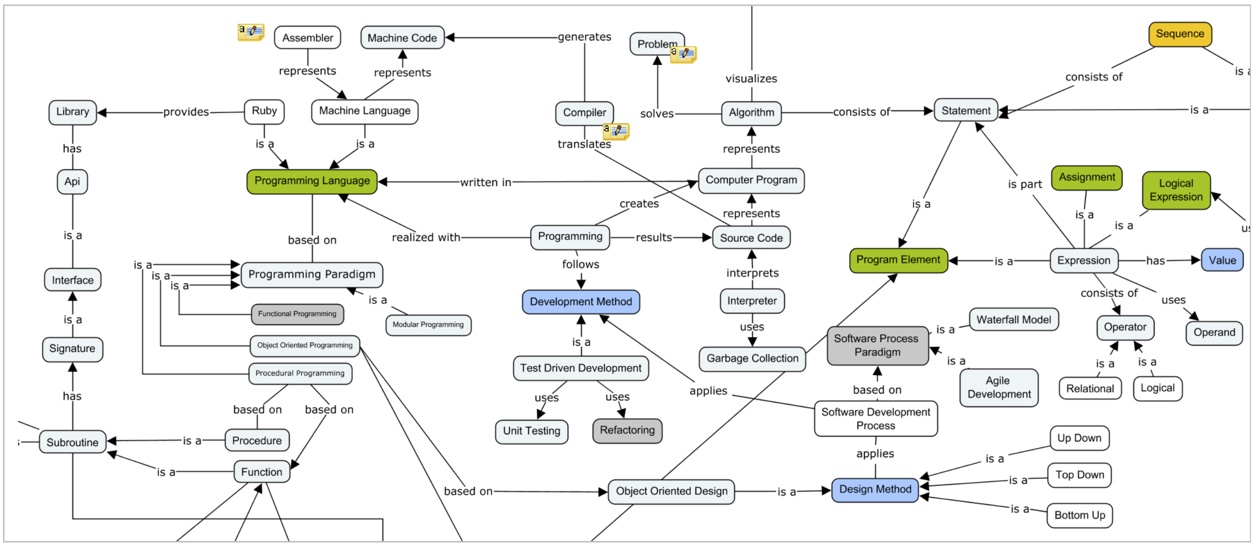

Figure 4: An extract of the CMap for software engineering.


```
<concept id="1RW00RMJG-16LJNQ6-35B" label="Procedure" short-comment="Prozedur" long-comment=
"Definition: Eine Prozedur ist ein Unterprogramm, das keinen Wert zur&#xfc;ckliefert, sondern
logisch zusammenh&#xe4;ngende Anweisungen zur Erf&#xfc;llung einer Aufgabe ausf&#xfc;hren"/>
<concept id="1YKL576XM-NM32F2-CNV" label="Software Process &#xd;Paradigm" short-comment=
"Softwareprozess-Paradigma" long-comment=""/>
<concept id="1RW03MVF1-VH8FDZ-4C1" label="Conditional" short-comment="Verzweigung"
long-comment="Synonyme (de): Auswahl, Selektion, Fallunterscheidung, Alternative&#xa;Related:
Bedingte Anweisung&#xa;&#xa;Synonyme (en): Branching, Choice, Selection&#xa;&#xa;Definition
(de):&#xa;Eine Verzweigung ist eine Kontrollstruktur, die abh&#xe4;ngig von einer oder
mehreren Bedingungen, einen bestimmten Programmabschnitt ausf&#xfc;hrt.&#xa;&#xa;Definition
(en):&#xa;Conditionals perform different computations or actions depending on whether a
condition evaluates true or false."/>
<concept id="1YJJLZM3Y-BHL9W5-K3S" label="Unit Testing" short-comment="UNIT_TESTING"
long-comment=""/>
<concept id="1YJJM5BQ1-26VF06V-L4J" label="Refactoring" short-comment="Refactoring"
long-comment=""/>
<concept id="1RW23TXR2-1H0W7QV-F1K" label="Access Rights" short-comment="Zugriffsrechte"
long-comment=""/>
<concept id="1YKL3Q2LY-16NC26Z-3GH" label="Top Down" short-comment="Top-Down" long-comment=""
/>
```


Figure 5: An example from the Cmap implementation, in an XML export

Our work consisted first in suggesting for the Cmap content an ontological representation framework, using RDF-based vocabularies for this. Figures 3 displays examples of such suggestions. We can also make use of the SKOS vocabulary for encoding properly the definitions, and we can use RDF(s) vocabulary for linking to other knowledge sources, like Wikidata, as this can be seen in the second example in Figure 6.

Our current work consists in using beyond SKOS a proper terminology representation model (TBX) in RDF, and in finalizing the integration of SKOS and our RDF-based representation of the Cmap data for having both resources really unified under one “onto-terminological” umbrella. We are also investigating the use of OntoLex-Lemon [3] (see https://www.w3.org/2016/05/ontolex/) for encoding the lexical semantics relations (synonymy and others), and thus establishing connection to lexicographic resources [4].

We hope to be soon able to demonstrate the use of such an integrated resource for the evaluative analysis of e-portfolios.

```
<http://example.org/aisop#1RW5G9MGQ-1XJJRH8-1HR>
 rdf:type owl:Class ;
 rdf:type skos:Concept ;
 aisop:hasID "1RW5G9MGQ-1XJJRH8-1HR" ;
 rdfs:label "\"Variable\""@de ;
 skos:exactMatch <http://example.org/aisop#1RW01GFY8-249X8S3-3N7> ;
 skos:inScheme aisop:FundamentaleIdeen-V2 ;
 skos:prefLabel "\"Variable\""@de ;
 skos:prefLabel "\"Variable\""@en ;
.
<http://example.org/aisop#1RW03SN8G-X86SZD-4R8>
 rdf:type owl:Class ;
 rdf:type skos:Concept ;
 aisop:hasID "1RW03SN8G-X86SZD-4R8" ;
 rdfs:label "\"Quellcode\""@de ;
 rdfs:label "\"code source\""@fr ;
 rdfs:label "\"source code\""@en ;
 rdfs:seeAlso <https://www.wikidata.org/wiki/Q128751> ;
 skos:inScheme aisop:Programmierung-V7 ;
 skosxl:altLabel aisop:label2-1RW03SN8G-X86SZD-4R8 ;
 skosxl:prefLabel aisop:label-1RW03SN8G-X86SZD-4R8 ;
.
```

Figure 6: Example of suggestions for the encoding of the PHW Cmap data in RDF, taking also multilingualism into consideration.

Summary

With this conceptual architecture description, we have provided a sketch of the systems that we expect to set-up in the AISOP project. The architecture allows us to see that the privacy concerns, which are of critical importance in the public education sector, can be satisfied with the current technologies. The architecture establishes itself in an eco-system of open-source systems and intends to be further implemented by open-source systems.

One of the biggest challenges of employing machine-learning in the processes is that making software available often also means making machine-learning models as well as the basis to generate these models, the corpora, available. Moreover, employing machine-learning tools requires annotation languages which can be encoded in a neutral fashion to be re-used in a diversity of tools. This is why the central annotation languages are studied for their translation to ontologies or terminologies, as we have done in the first strokes of the knowledge architecture.

References

[1] Wolfgang Mueller, Sandra Rebholz, and Paul Libbrecht. Automatic inspection of e-portfolios for improving formative and summative assessment. In Ting-Ting Wu, Rosella Gennari, Yueh-Min Huang, Haoran Xie, and Yiwei Cao, editors, Emerging Technologies for Education, pages 480–489, Cham, 2017. Springer International Publishing.

[2] Christophe Roche. Ontoterminology: How to unify terminology and ontology into a single paradigm. In International Conference on Language Resources and Evaluation, 2012.

[3] John P. McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, Philipp Cimiano. The OntoLex-Lemon Model: Development and Applications. Proceedings of eLex 2017

[4] Thierry Declerck and Melanie Siegel. OntoLex as a possible Bridge between WordNets and full lexical Descriptions. Proceedings of the Global WordNet Conference 2019