EPiC Series in Computing
Volume 105, 2025, Pages 314–325
Proceedings of EUNIS 2024 annual congress in Athens

Artificial Intelligence Enhanced Course Content Labelling

Conceptual Model and Testing with University Teachers

Ville Kivimäki[1], Sami Hautakangas[2], Heli Järvelä[1] and Patrik Maltusch[1]

[1] Aalto University, Espoo, Finland
[2] Tampere University, Tampere, Finland

ville.kivimaki@aalto.fi, sami.hautakangas@tuni.fi,
heli.jarvela@aalto.fi, patrik.maltusch@aalto.fi

Abstract

This study presents a novel approach by investigating the efficacy of an artificial intelligence (AI) -aided tool, Annif, in generating course-specific keywords for university courses. With the increasing number of university courses, there is a need for an effective method to assist students in navigating and selecting courses based on their interests and job market compatibility. The traditional manual keyword approach, while accurate, can be laborious and time-consuming. AI has the potential to automate this process, but it is crucial for teachers to validate the results to ensure accuracy. This study seeks to explore the potential of AI in this context, addressing two key questions: the ability of AI-generated keywords to establish course connections and the reactions of teachers to the use of AI in keyword generation. The results reveal that the AI tool can provide accurate keywords for about 64% of the courses. While teachers found this approach useful, the study highlights the need for teacher validation to ensure the accuracy and appropriateness of the AI-generated data. Therefore, while AI can significantly contribute to keyword generation, human intervention is still indispensable to maintain its quality in the academic context.

1 Introduction

The massification of education has increased the number of universities, courses, and students (Evans et al., 2021). The widening offering of courses and course topics generates demand for ways to monitor the course offerings, help students find courses that meet their interests and navigate between the courses (Ota & Mima, 2011). There has been an addressed need to help learners compare their skills and competencies with the opportunities in the labour market. European Commission has developed ESCO, a classification for skills, competencies, qualifications and occupations for this purpose (European Commission, 2013). In addition, there can be a national need to start collecting metadata, such as course content keywords. In Finnish Digivisio 2030, the metadata is needed to help learners find education offerings that meet their situation, schedule, and competence needs (Digivisio 2030 The Digivisio 2030 programme, 2022). One partial solution here would be to classify the courses by their content, e.g., by requesting teachers add course-specific keywords to the course syllabus and use that metadata to retrieve the relevant courses later by using them in search inquiries (Golub et al., 2016).

The manual keyword approach has two main problems. First, course content classification can be labour-intensive work with thousands of course descriptions even if content specialists, such as teachers, know their course content thoroughly. Second, the findability and linking between the courses would require using agreed vocabulary. Large and complex vocabularies add to the labour-intensiveness. For example, ESCO vocabulary for skills and competencies contains more than 14,000 terms. The best solution to diminish the labour of generating keywords would be to automatise the process fully. Conversely, the automatically generated keyword could be wrong (does not fit the course content), too general (`history´), too specific (`hydrogen´ on general chemistry course), or even dangerous (e.g., politically incorrect or defamatory/coarse language). This could require the exploration of semi-automated solutions where teachers could play a role in validating and, when needed, editing the results. It is not yet known how any of the scenarios would work out. The accuracy of the current state-of-the-art indexing is yet to be known. In addition, we do not know how teachers would react to a semi-automatised approach.

Teachers are not the first ones to benefit from course keyword generation. The new metadata would first benefit the organisational aims and serve students. Certainly, it can be argued that teachers would benefit from finding other courses and teachers with similar course contents. Still, teachers would be the ones to consider when designing the course keyword generation and collection procedure. Thus, this paper focuses on teachers and how we could automatise or semi-automatise the keyword generation, minimising the effort required from individual teachers.

The current study focuses on building an automated algorithm-based course keyword generation solution in the context of university courses and aims to answer the following research questions:

RQ 1: Do the generated keywords offer opportunities for connecting courses?

RQ 2: How do teachers rate the keywords matching their course content?

In addition, this study builds a conceptual base for the practice. Here, we aim to expand the concept to other service products and areas in the context of university education. This aim stems from the practice that solutions that solve or advance various business demands are desirable in the context of limited resources.

2 Method

2.1 Tools and data

Keywords in this study are from YSO - General Finnish Ontology, maintained by the National Library of Finland (*YSO - General Finnish Ontology*, 2024). The ontology has been designed primarily to be used by libraries, archives, museums, and other content providers for describing, categorising, and retrieving data. YSO is a trilingual ontology with Finnish, Swedish, and English concepts. It includes concepts from a broad range of topics and categories. YSO consists of concepts identified with URIs and associated labels in three languages. Each concept is placed in a hierarchical structure, guided by the principles of the Simple Knowledge Organization System (SKOS) model. For instance, if "fruit" were a concept present in the ontology, "apple," "banana," and "orange" might be some of its narrower concepts. Moreover, the relationships between concepts aren't limited only to hierarchical aspects. YSO also incorporates associative relationships, which connect related concepts that aren't in a hierarchical relationship. Throughout its structure, every concept and relationship in YO is equipped with detailed annotations, including definitions, scope notes, and change notes. This extensive metadata is designed to provide better context and clarity for users.

For subject classification, we utilised an open-source tool and microservice, Annif (*Annif - Tool for automated subject indexing and classification*, 2023). The service uses several algorithms to suggest a list of keywords based on the natural language input (Suominen, 2019). The latest version accommodates a large language model based algorithm on top of several machine learning algorithms (Suominen et al., 2022). The service is available in three languages: Finnish, Swedish, and English. This implies that every concept described in the YSO ontology has an equivalent expression in each of these languages. The API also supports all three languages, but defining the language is necessary while making an API call.

Data processing procedure in Figure 1 depicts how course content data sent for the Annif service for automated indexing was connected with the keyword suggestions from Annif and visualised with Power BI. The course data has several dimensions (course name, department, teaching period, etc.).

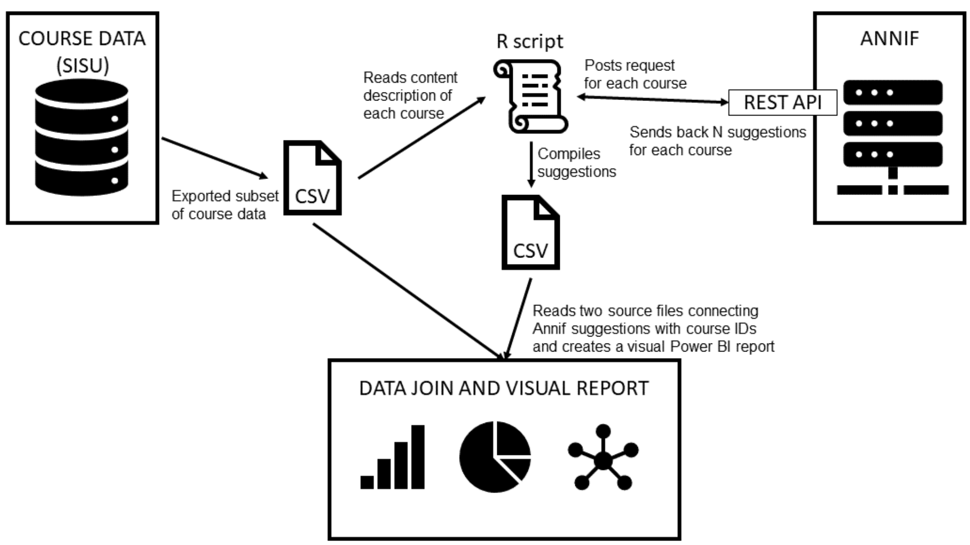

Figure 1: Modular architecture of data flows in the experimental POC setup

2.2 Educational context

Four Finnish universities started a proof-of-concept project in August 2023 to learn how the AI-based classification would technically work with their student information service Sisu[1] and to find out

[1] Joint effort of originally four Finnish universities to modernize student registry system. Now a consortium of 8 Finnish universities and universities of applied sciences.

how the generated keywords would support the academic staff in validating the metadata as a part of the curriculum planning process.

Tampere University has started preliminary tests with academic faculties to see how the automatically generated keywords fit the content of the course units described in the curriculum. The descriptions of the names, learning outcomes, and content of the course units were produced as a CSV file from the Sisu system, and the Annif results were combined with the course data. The empirical testing is still in the early stages. The testing sessions have been organised as follows.

The specialists supporting the curriculum planning related to the new needs of continuous learning gathered information on teachers interested in piloting the development of the studies offered in continuous learning. AI-supported classification of the curriculum data has been only one part of this piloting. The 14 teachers who have participated in the tests so far have been from various fields of study (mathematics, computer science, information technology, chemistry, history, social sciences, psychology and social work). The teachers are experienced academic staff who are responsible for the course units that were part of the testing and are very capable of evaluating the fitness of the terms to the content of the course units.

The Excel sheets were prepared for the teachers, where Annif labelled each course unit (n=26) with ten terms with the highest probability. For data evaluation, there are columns for simple fit-for-purpose evaluation (1=fits the purpose, 0= does not fit the purpose), comments on the suggested term, and possible alternative suggestions from the teacher. The testing phase started with an introduction session, where the concept of AI-based classification was presented, and certain examples were discussed with the teachers. The teachers evaluated the terms using Excel sheets. To facilitate the work and to get direct feedback from the teachers, there have been three different workshops by the 10th of April, and there will be more for the teachers from other faculties. In the workshop, the teachers were advised to suggest more suitable terms to substitute for the failed ones, where YSO – General Finnish ontology was given as support. The findings of teachers were discussed at the end of each workshop. The results from this preliminary phase are presented in the next chapter.

The curriculum design process (Figure 2) In the university environment, curriculum analysis and development usually comprise curriculum analysis and development, gathering curriculum data, decision-making, and publishing of the curriculum. In the curriculum analysis and development, the programme director, teachers, and learning services study the feedback from the students, faculty and stakeholders and summarise the needed changes to the curriculum. The teachers gather the changes directly in the student information system or specified application designed for this purpose or manually with the help of Excel spreadsheets and other working documents. The academic committee decides the curriculum, and after the decision, it is published for students in the student information system and the university's website.

In the analysis phase of curriculum development, the need for a mapping tool has also risen at Aalto University. This has been done in a very laborious manner with large Excel spreadsheets for a degree programme. The responsible teachers of the courses have marked themselves as to whether the course they are teaching enhances a specific learning objective of the degree programme. For this purpose, the help of an agreed vocabulary and AI-based tool for teachers to search for suitable keywords would be greatly appreciated.

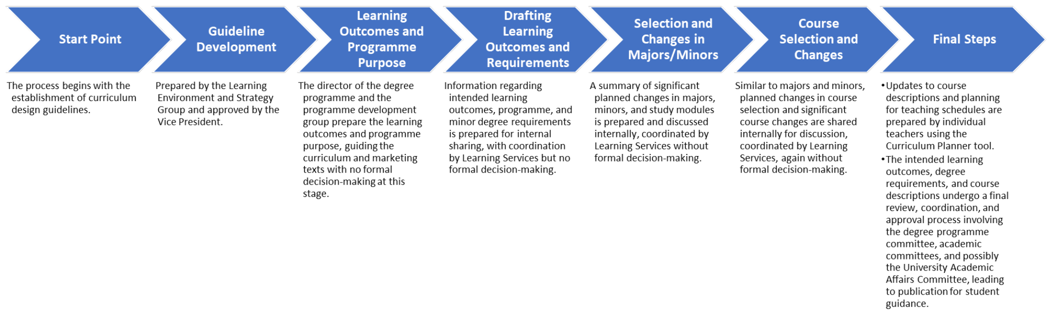

Figure 2: Curriculum design practice

3 Results

3.1 AI-based keyword generation for analytics

Analytics opportunities were tested with a large set of courses. Data consisted of course content descriptions that were written in English. A total of 3 616 course descriptions were sent to the Annif service, which resulted in 36 124 keywords (3 504 unique). Each course was expected to be awarded ten keywords (36 160). Hence, the API connection missed some data (0.1%) for technical capacity reasons in the API connection. The missing data was not relevant to the accuracy of this study. The results included the keyword, the scores and the link to the keyword's ontology page (Table 1).

Course name	keyword	Score	Uri
Parametric CAD	CADS	0.11371873319149	http://www.yso.fi/onto/yso/p20149
Parametric CAD	computer programmes	0.0552481710910797	http://www.yso.fi/onto/yso/p26592
Parametric CAD	computer-aided design	0.45993047952652	http://www.yso.fi/onto/yso/p17865
Parametric CAD	design (artistic creation)	0.113796919584274	http://www.yso.fi/onto/yso/p6455
Parametric CAD	drawing (artistic creation)	0.190325289964676	http://www.yso.fi/onto/yso/p695
Parametric CAD	modelling (representation)	0.134560152888298	http://www.yso.fi/onto/yso/p3533
Parametric CAD	PC Tools Utilities	0.0861699730157852	http://www.yso.fi/onto/yso/p13301
Parametric CAD	planning and design	0.0632747039198875	http://www.yso.fi/onto/yso/p1377
Parametric CAD	students	0.177192032337189	http://www.yso.fi/onto/yso/p16486
Parametric CAD	three-dimensionality	0.096014067530632	http://www.yso.fi/onto/yso/p1978

In total, 1,396 keywords are uniquely assigned to only one course. Hence, most (over 60%) keywords are assigned to two or more courses. Moreover, as courses are assigned ten keywords each, several connections can be found between the courses (Figure 3).

Figure 3: Network visualisation between the courses (black dots) connected by the keywords (grey dots).

The network visualisation revealed, for example, the high-density cluster of similar courses (maturity test) at the bottom of Figure 3.

Keywords were generated using two different sources: only the course content description and course content description were added with learning outcomes. The School of Electrical Engineering courses yielded results in Figure 4 and Figure 5.

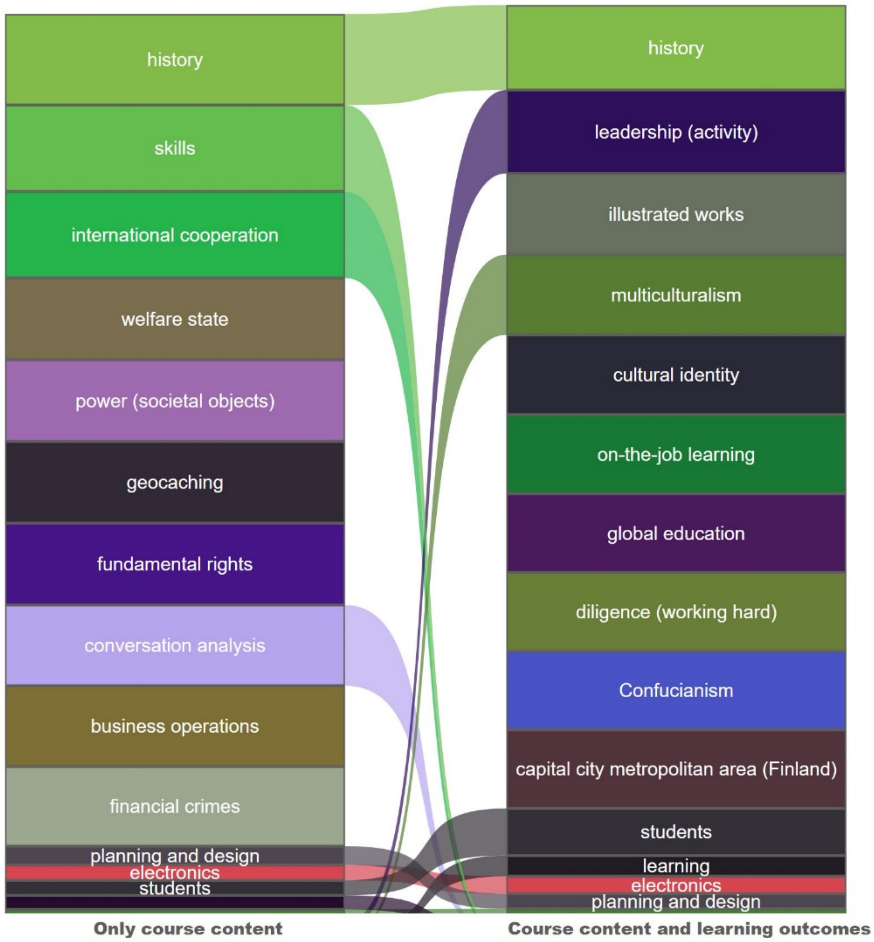

Figure 4: Keywords on the electrical engineering field courses sorted by the number of courses with the same keyword and comparisons between two keyword source data methods.

Figure 5: Word cloud of electrical engineering field course keywords

3.2 Teacher validation of keywords

A validation study was conducted using a different dataset, where some courses had descriptions only in Finnish while others had descriptions only in English. Finnish language was given priority, meaning Finnish keywords were assigned if a course had a Finnish content description. If there was no Finnish content description, the English content was used to generate keywords in English, which were then assigned to the courses that still lacked keywords.

The results from the first preliminary tests from Tampere University already give some guidelines for the next steps of the development. Even if the sample size is small, the qualitative aspects from the evaluation and the following discussion make certain aspects of the tested concept quite clear. Firstly, from the 260 keywords generated for the 26 course units by Annif, 138 keywords were evaluated to fit the content of the course. The accuracy percentage of this experiment was thus 53%. The accuracy percentage varied between course units from 20% to 70%. Overall, the teachers' comments in the discussion were typically along the lines of "it was ok" and "it seemed helpful", and they were willing to continue the work. From this result, we can draw one important conclusion. Even if Annif is trained well for its purpose, *the accuracy of AI-based classification is too low to be relied on as the only method.* This is especially important when the curriculum metadata needs to be as valid as textual descriptions. There have been ideas that the curriculum data could be enriched using AI after the publication of the official curriculum, but this finding does not support the idea. Therefore, the validation of the keywords must be integrated into the curriculum planning process so that the teachers or other experts who are best able to evaluate the suitability of the terms can do this before the curriculum is ready.

The second part of the results consists of teachers' comments. Most failed terms were deemed "overly general" for the purpose. Very few terms were wrong, but they were missing what was essential in the course. Some more specific terms were "not quite what the course was about". An important aspect that affects the accuracy is how the textual descriptions of learning outcomes and content are written because even in the small number of course units, there was considerable variation in the level of detail. The style in which the learning outcomes are written will easily lead to emphasising certain general terms of education because some expressions are often repeated (“After completing the course, the student will...”). There was also discussion that some of the terms were redundant with other terms. This discussion indicated that the suggested terms had only a relatively small number of total misses. It was quite straightforward for the teachers to suggest better-fitting terms to fix the ones that were not suitable. One finding was that even if YSO ontology contains more than 30,000 terms, there was a lack of suitable terms in certain specific areas. The National Library of Finland maintains YSO ontology and has a process for suggesting new terms so that new needs can be addressed by using the existing service.

The overall experience of the data produced by Annif was that it was not totally amiss. This explains why the teachers had a positive attitude towards this classification method as part of the curriculum planning. One should note that there is also a psychological effect when the academic ownership of teachers on the content of the courses is respected and appreciated even if the part of the process involves “only metadata”. There are also possibilities to improve the accuracy of suggestions by utilising the evaluation data from teachers to train the Annif further. If the evaluation can be linked to curriculum planning processes, the data from one university alone can cover thousands of course units.

To develop the concept and the tools supporting it, teachers would need support to substitute the “not-so-good" terms with more appropriate ones. This would mean, for example, an interface to Annif to enable the showing of the more specific (or more general) terms in the YSO hierarchy, where the teachers could select better terms. And of course, because the first experiment was arranged by using an Excel sheet, there were expectations that the features would be integrated into the actual tool the teachers are using in the curriculum planning, either as a feature of the Sisu system or as a plug-in working seamlessly with Sisu.

4 Discussion

Research question 1 explored the analytics capabilities related to keyword generation. The descriptive analytics drawn from the keywords clearly show promise in connecting courses with matching keywords. Still, nearly 40% of all the keywords were only assigned to one course. One possible use case for connected courses is a course catalogue, where users could use keywords to navigate between courses. Although 60% of the keywords connected several courses, thus allowing this use-case, nearly half of the keywords would not lead to any other course. This could be overcome by highlighting the keywords that lead to other courses or using semantic matching techniques on top of the keyword approach. A more robust approach could be to ask for more than ten keywords and force the course catalogue to list only the top N keywords that connect to other courses.

The browsability advancement cases are one possible and desirable outcome of the metadata generation with keywords. Perhaps an even more obvious case is to describe the large pool of courses, e.g., at the school, faculty or department level. Here, the word clouds could come in handy, offering also browsability options on reports made with business intelligence tools, e.g., Power BI in Figure 5.

Research question 2 studied teachers’ validation of their own course’s keywords, i.e., how well the AI-generated keywords match the course content. Based on the teacher validation results, the accuracy of the keyword validation has generated a positive outlook. Preliminary results suggest 53% accuracy, which is close to earlier results in the most recent literature (Suominen et al., 2022), a clear advance from the first validation results just a few years earlier (Suominen, 2019). Nevertheless, the results are still poor enough to prevent the possibility of fully automatic keyword generation and publication. As described in the method section, a university context with a rigorous curriculum process plays an integral role here. Even the AI-generated keywords could be interpreted as official curriculum data, based on which students can make study planning decisions. Fallacious keywords could generate liability risks as the university is always accountable for the information it produces and shares. When utilising AI, it is important to take into account its purpose. The requirements for the data are very different when the data is used, e.g. for automated profiling or decision-making, compared to more modest uses.

Technically, the solution presented in this study lays a foundation for multiple use cases of such data. The AI-generated keywords by Annif service are reasonably accurate and positively surprise even the experts from specific disciplines. Still, human validation is needed, mainly when handling official data. When connected with human validation, e.g., as a part of the curriculum decision process, the Annif could ease the teacher's workload. Elsewhere, the keywords could be used more freely. One case is analytics. Connecting courses with matching keywords offers high-level web usability, and word clouds could help decision-makers get a fast glimpse of the course contents. This metadata could also help teachers find other courses, and consequently, other teachers, that are teaching similar content. Certainly, the same keyword could have several meanings in different fields of study. Even if this is the case, fruitful discussions could happen.

The 50-70% accuracy level, at best, as found in our study and literature, offers a good foundation to work with. That said, this applies to one keyword. Other ontologies, such as ESCO, combining most keywords with verbs, could result in surprisingly low accuracy levels. To illustrate this, imagine AI generating a verb with an accuracy of 70% and a related keyword with an accuracy of 70%. The probability of hitting it right lowers to 49% (0.7 * 0.7 = 0.49). And to think that with courses where only two out of ten keywords match the course content, such service would, at best, be unavailing. From another angle, the accuracy of AI-based keyword generation relies a lot on human contextual understanding. Still, several other than curriculum use contexts can be imagined at universities.

- Website keyword generation
- Research project (website) keywords
- People profile finder
- Research paper keyword generator
- Decision document repository's searchability

To our knowledge, the Annif service is currently already being used to index student's theses (*Annif - Tool for automated subject indexing and classification*, 2023). Future studies could involve testing Annif against general-purpose AI models, such as Llama 2, GPT 3.5, GPT 4.0, and Gemini. The hypothesis could be that the vocabulary would extend uncontrollably, thus reducing analytical power compared to Annif with a stable and curated ontology. However, this problem could be fixed with advanced prompts, e.g.

In sum, the recent development in artificial intelligence shows much promise even when validated by field experts. These results are not the fruits of a general AI but a highly specialised service, which could still be fine-tuned to fit each purpose. Still, it is clear that while human validation and processing are still needed in demanding environments, AI services are creating new possibilities for natural language processing in analytics and helping in metadata generation, a much-needed feature of development for future services.

References

Annif - Tool for automated subject indexing and classification. (2023). National Library of Finland. Retrieved 30/01/2024 from https://www.annif.org

European Commission, D.-G. f. E., Social Affairs and Inclusion. (2013). *ESCO, European classification of skills/competences, qualifications and occupations: the first public release: a Europe 2020 initiative*. Publications Office.

Evans, C., Rees, G., Taylor, C., & Fox, S. (2021). A liberal higher education for all? The massification of higher education and its implications for graduates' participation in civil society. *Higher education*, *81*(3), 521-535. https://doi.org/10.1007/s10734-020-00554-x

Golub, K., Soergel, D., Buchanan, G., Tudhope, D., Lykke, M., & Hiom, D. (2016). A framework for evaluating automatic indexing or classification in the context of retrieval. *Journal of the Association for Information Science and Technology*, *67*(1), 3-16. https://doi.org/https://doi.org/10.1002/asi.23600

Ota, S., & Mima, H. (2011). Machine Learning-based Syllabus Classification toward Automatic Organization of Issue-oriented Interdisciplinary Curricula. *Procedia, social and behavioral sciences*, *27*, 241-247. https://doi.org/10.1016/j.sbspro.2011.10.604

Suominen, O. (2019). Annif: DIY automated subject indexing using multiple algorithms. *LIBER quarterly*, *29*(1), 1-25. https://doi.org/10.18352/lq.10285

Suominen, O., Inkinen, J., & Lehtinen, M. (2022). 'Annif and Finto AI: Developing and Implementing Automated Subject Indexing', 'Annif and Finto AI: Developing and Implementing Automated Subject Indexing'. *JLIS.it : Italian journal of library and information science*, *13*(1). https://doi.org/10.4403/jlis.it-12740

YSO - General Finnish ontology. (2024). National Library of Finland. Retrieved 10/02/2024 from https://finto.fi/yso/en/

6 Author biographies

Analytics Technology Owner **Ville Kivimäki** is working as an IT Specialist at Aalto University. He was the project manager in an online learning pilot that resulted in a learning analytics product called Course Diaries. He has worked at Tampere University, Aalto University and the University of Helsinki as a research curator, teacher, Planning Officer, Team Leader, Project Manager, Data Science Specialist, and Doctoral Researcher. Work topics have varied around curriculum work, pedagogical development, student success, and retention. He has authored more than ten scientific publications.

LinkedIN: www.linkedin.com/in/ville-kivimaki; Research profile: https://researchportal.helsinki.fi/en/persons/ville-antero-kivim%C3%A4ki.

Sami Hautakangas works as a Senior Specialist in Services for Educational Leadership at Tampere University. He has 20+ years of experience in developing higher education and the digital support for processes. Sami Hautakangas has experience in several national cooperation projects of HEIs in Finland and has participated actively in the development of national specifications promoting interoperability between the HEIs' information systems.

Digital Learning Services Solution Owner **Heli Järvelä** is working as an IT Specialist at Aalto University. She has previously worked as a Study Coordinator, Planning Officer and Specialist in Study Information Services at Aalto University.

LinkedIN: www.linkedin.com/in/heli-järvelä-51010b4

Patrik Maltusch is the head of EA architecture tea at Aalto University. He is chairing the Finnish EA-SIG, EUNIS EA-SIG and has been one of the lead educators coaching administration staff in the national Higher Education EA program. Early experience, include working as a customer service instructor and further fifteen years as network architect and business owner for infrastructure design in a global Telco company. Patrik is also a distinguished and accredited security professional, risk manager, system auditor and Education Enterprise Architect. For Patrik interoperability is the key to master the ever-growing digitalization needs in a more complex and complicated ecosystem landscape.

LinkedIN: https://www.linkedin.com/in/maltusch