

Automating Subject Indexing at ZBW: Making Research Results Stick in Practice

Anna Kasprzik

ZBW – Leibniz Information Centre for Economics, Hamburg/Kiel, Germany,
a.kasprzik@zbw-online.eu, orcid.org/0000-0002-1019-3606

Abstract

Subject indexing, i.e., the enrichment of metadata records for textual resources with descriptors from a controlled vocabulary, is one of the core activities of libraries. Due to the proliferation of digital documents, it is no longer possible to annotate every single document intellectually, which is why we need to explore the potentials of automation on every level.

At ZBW the efforts to partially or completely automate the subject indexing process started as early as 2000 with experiments involving external partners and commercial software. The conclusion of that first exploratory period was that commercial, supposedly shelf-ready solutions would not suffice to cover the requirements of the library. In 2014 the decision was made to start doing the necessary applied research in-house which was successfully implemented by establishing a PhD position. However, the prototypical machine learning solutions that they developed over the following years were yet to be integrated into productive operations at the library. Therefore in 2020 an additional position for a software engineer was established and a pilot phase was initiated (planned to last until 2024) with the goal to complete the transfer of our solutions into practice by building a suitable software architecture that allows for real-time subject indexing with our trained models and the integration thereof into the other metadata workflows at ZBW.

In this paper we address the question of how to transfer results from applied research into a productive service, and we report on the milestones we have reached so far and on those that are yet to be reached on an operational level. We also discuss the challenges we were facing on a strategic level, the measures and resources (computing power, software, personnel) that were

needed in order to be able to affect the transfer, and those that will be necessary in order to subsequently ensure the continued availability of the architecture and to enable a continuous development during running operations. We conclude that there are still no shelf-ready open source systems for the automation of subject indexing – existing software has to be adapted and maintained continuously which requires various forms of expertise. However, the task of automation is here to stay, and librarians are witnessing the dawn of a new era where subject indexing is done at least in part by machines, and the respective roles of machines and human experts may shift even further and more rapidly in a not-so-distant future. We argue that in general, the format of “project” and the mindset that goes with it may not suffice to secure the commitment that an institution and its decision-makers and the library community as a whole will have to bring to the table in order to face the monumental task of the digital transformation and automation in the long run. We also highlight the importance of all parties – applied researchers, software engineers, stakeholders – staying involved and continuously communicating requirements and issues back and forth in order to successfully create and establish a productive service that is suitable and equipped for operation.

Keywords: subject indexing; automation; machine learning; artificial intelligence; metadata; IT infrastructure

1. Introduction – Context

Subject indexing, i.e., the semantic enrichment of metadata records with descriptors, is one of the core activities of libraries. Trivially, due to the proliferation of digital documents it is no longer possible to annotate every single document intellectually – which is why we need to explore the potentials of automation on every level. Automation can start with small measures such as using simple scripts and routines for metadata manipulation and go all the way to the use of methods from Artificial Intelligence, notably from the domain of Machine Learning.

At ZBW the efforts to automate the subject indexing process started as early as 2000. Two projects with external partners and/or commercial software yielded some insights into the state of the art at the time but mostly showed that the evaluated solutions would not suffice to cover the requirements of the library and that there still were several hurdles to overcome both with

respect to the quality of the output and to the technical implementation. However, abandoning the endeavour was not an option since the need for automation became ever more obvious and pressing over time. A reorientation phase around 2014 led to the decision that from then on, the necessary applied research should be done in-house and only open source software should be used and created. For this purpose, a full-time position for a research engineer with the option to obtain a PhD in computer science was established within the library. The first phase of activities after this reorientation was called “project AutoIndex” and lasted until 2018. After a personnel change in 2018 the role of coordinating the automation of subject indexing was upgraded to a permanent full-time position and filled with a computer scientist with additional library training.¹

However, the prototypical machine learning solutions that were developed in project AutoIndex were not yet ready to be integrated into productive operations at the library. In order to be able to take on this challenge properly, several additional adjustments were made on the strategic level: Most importantly, the automation of subject indexing at ZBW was declared no longer a project but a permanent task (dubbed “AutoSE”). This in turn prompted the initiation of a pilot phase (starting in 2020, planned to last until 2024) with the goal to transfer results from applied research in the AutoSE context into a productive service by building a suitable software architecture that allowed for real-time subject indexing with the trained AutoSE models and integration thereof into the other metadata workflows at ZBW. In order to meet these requirements, AutoSE was granted one more full-time position: since the beginning of the pilot phase, the team consists of a staff of three, covering the roles of lead/coordination, applied research, and software development/architecture.

2. Applied Research and Productive Operations

2.1. Applied Research – Methods

From the machine learning point of view, subject indexing is a so-called multi-label classification task, i.e., to each publication several labels (~subjects) can be assigned. Since the end of the last AI winter (around 2012) more and more – actually usable! – machine learning models for this task have emerged, and a

large portion of them are available as open source software. In the precursor project, AutoIndex, a prototypical fusion approach towards automated subject indexing at ZBW had been developed that joined several methods and then filtered their combined output using additional rules (Toepfer & Seifert, 2018a). At the same time, a team at the National Library of Finland (NLF) started creating the open source toolkit Annif (Suominen et al., 2023a) which offers various machine learning models for automated subject indexing and also allows the integration of one's own models. The two institutions were in contact and exchanged information about their respective developments.

At the beginning of the pilot phase the AutoSE team adopted Annif as a framework in order to combine several state-of-the-art models – currently the following four are used: two variants (*parabel* and *bonsai*) of *omikuji* (Dong & Suominen, 2022),² which are tree-based machine learning algorithms, *fastText* (Facebook Inc., 2022; Joulin et al., 2016),³ which uses word embeddings, and *stufsa* (ZBW, 2022a),⁴ a lexical algorithm based on finite-state automata, which was developed at ZBW and is optimised for the “Standard-Thesaurus Wirtschaft” (STW) (ZBW, 2023), the thesaurus for the economics domain hosted and used for subject indexing at ZBW, but can be used with other vocabularies as well. The output of all of these methods is then combined via another method – *nn-ensemble* (Suominen et al., 2023b) – which balances them out, yielding as final result a set of subjects that have all passed a given confidence threshold. For AutoSE the models are trained with short texts from the metadata records underlying the ZBW research portal EconBiz (<https://www.econbiz.de/>), specifically titles and (if available) author keywords⁵ of publications in English. In parallel, applied research continues in order to explore other machine learning methods beyond the classical ones, including approaches from Deep Learning in the form of large language transformer models, notably pretrained ones (GPT (Radford et al., 2018) is a prominent example), which are particularly promising for multi-lingual subject indexing (“Transformer (machine learning model)”, 2023).

The AutoSE team is actively involved in the continuous advancement of Annif, checking with NLF at regular intervals if results from the AutoSE context can be integrated as new functionalities, assisting NLF with giving tutorials, and other institutions with advice on how to deploy Annif in practice.⁶ The team has complemented the ZBW instance of Annif with their own components for setting up experiments, hyperparameter optimisation, various additional quality control mechanisms (see Section 3),

and APIs in order to communicate with internal and external metadata workflows.

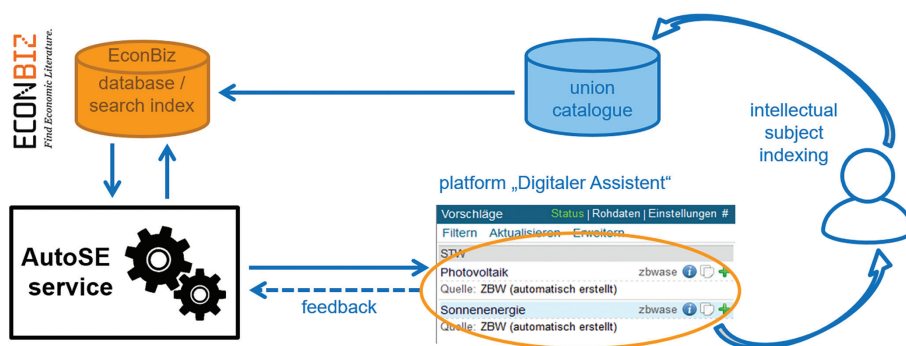
2.2. Productive Operations – Data Flows

A first version of a productive AutoSE service went into operation in 2021. The software runs on a Kubernetes cluster of five virtual machines and technologies such as *helm*, *GitLab*, *prometheus* and *grafana* are used for software deployment, continuous integration, and monitoring. As the applied research continues and the team is integrating more and more of the original requirements as well as supplementary enhancements, the architecture is constantly evolving and its modular design keeps it adaptable to future developments beyond the pilot phase.

The output of the service is used for two purposes at present: The first is fully automated subject indexing, for publications in English that would otherwise not be annotated with any subjects from the STW thesaurus at all. The EconBiz database is checked every hour for new eligible metadata records, these are then enriched by AutoSE with STW subjects, and written back into the database immediately. If a publication happens to belong to the core set of literature that is earmarked to be annotated by human specialists at ZBW then the AutoSE subjects are subsequently suppressed both in the search index and in the single display page for this publication once the intellectual subject indexing has taken place. The connection between AutoSE and the EconBiz database was activated in July 2021, and in the first six months of operations, over 100,000 machine-annotated metadata records were entered into the database via direct write access.⁷ The total number of records enriched by AutoSE methods in the database is higher as the team also processes large amounts of records retroactively which are then written back into the database via a batch process.⁸ As of December 2022, the EconBiz database contains around 1.3 Mio. records with AutoSE subject indexing, which corresponds to about a quarter of ZBW holdings.

The second purpose of the output of the service is machine-assisted subject indexing: the subjects generated by AutoSE are made available as suggestions to the platform used for intellectual subject indexing at ZBW (“Digitaler Assistent”; DA-3)⁹ via an API. This connection was the first one implemented, in 2020. Within DA-3, AutoSE suggestions are marked as machine-generated

Fig. 1: Data flows of machine-generated subject indexing using the AutoSE service.



for reasons of transparency, and they can be adopted by a single click on an “add” button during the annotation of a publication. Freshly annotated records are stored in the union catalogue and mirrored back into the EconBiz database where the AutoSE team collects them and computes the F1 score (“F-score”, 2023) from the difference in order to monitor the performance of the current productive backend (also see Section 3). Figure 1 shows an overview of the corresponding data flows.

Milestones yet to be completed within the pilot phase include:

- Preparing for the use of abstracts and potentially also tables of content in addition to titles and author keywords – besides gathering the necessary amount of training data for experiments in order to develop models that are optimised for these kinds of text materials, the team also had to clarify the situation with respect to text mining rights for abstracts since most licences do not mention the use of abstracts for non-commercial productive purposes such as the AutoSE service, even if the use for research purposes is explicitly allowed.
- Preparing the integration of solutions for languages other than English – approaches include an upstream machine translation before subject indexing and, as the most promising option, the use of BERT-like transformer models (“Transformer (machine learning model)”, 2023).
- Finalising and publishing a web user interface which provides an interactive demo of the productive backend, statistics concerning the

current and past performance of the AutoSE service and additional information about the methods comprised in the backend.

- Automating various machine learning processes such as hyperparameter optimisation and training, in order to be able to retrain the models more easily when enough new metadata records have accumulated or a new version of the STW thesaurus is available.
- Documenting the requirements of (future) productive operations (also see Section 4).

Plans beyond the pilot phase include extending the architecture to integrate automated metadata extraction workflows in order to generate more input for AutoSE, and combining machine learning with symbolic approaches, i.e., incorporating more semantic information from STW and from external sources as a way to check the plausibility of the output of our trained models.

3. Quality Management

3.1. Various Approaches to a Quality Assurance Concept

The automation of subject indexing is a change prompted by new technological possibilities, but it also affects subject indexing practices on a cultural level. In an automation endeavour such as this, quality control is key – both because of the (positive or negative) effects of metadata quality on retrieval and because the approval of the output of the service among the stakeholders (i.e., in particular subject indexing experts) is vital in order to make sure that it will be accepted and used long-term.

The AutoSE team is working on a comprehensive quality assurance concept using different approaches in order to be able to guarantee an overall subject indexing quality that is as high as possible (also see Kasprzik (2022)). On the technical side this includes working with metrics commonly used in the machine learning domain (currently aiming to maximize the F1 score but we plan to evaluate differently weighted combinations of precision and recall or other metrics such as Normalized Discounted Cumulative Gain or metrics that take the hierarchy of the thesaurus into account as well) and identifying reasonable thresholds (e.g., the minimum level of confidence required). After the automated subject indexing process proper, those thresholds are applied to the output (along with other filters and if-then-rules such as blacklists and

mappings, see 3.2). Since 2022, quality control for AutoSE also features the application of a machine-learning-based approach for the prediction of overall subject indexing quality for a given metadata record. More precisely, the method *qualle* predicts the recall for that record by drawing on confidence scores for individual subjects plus additional heuristics such as text length, special characters, and a comparison of the expected number of labels to the number of labels that were actually suggested. *Qualle* is based on a prototype described in Toepfer and Seifert (2018b) – however, in order to be usable in productive operations, the code had to be re-implemented from scratch (ZBW, 2022b). Before launching *qualle* the team asked ZBW subject indexing experts for an intellectual review of its output in order to make sure that it would outperform the previous method – up until that point a much coarser semantic filter had been applied for quality control on the metadata record level (rule “min2VB”): the output had to contain at least two subjects from one of the two economic core domains, modelled as two sub-thesauri in STW. Thus, this heuristic neglects other domains associated with economics, whereas *qualle* simply learns from the training data what an appropriate subject indexing should look like, without discriminating between sub-thesauri. This shows that, if trained on suitable data, a machine-learning-based method can be more flexible than an intellectually postulated rule.

3.2. The “Human in the Loop”

However, one of the most essential components of quality assurance is and will remain the human element. The machine learning domain has coined the phrase “human in the loop” for this paradigm, which addresses “the right ways for humans and machine learning algorithms to interact to solve problems” (Monarch & Manning, 2021). Possible interpretations and implementations may include:

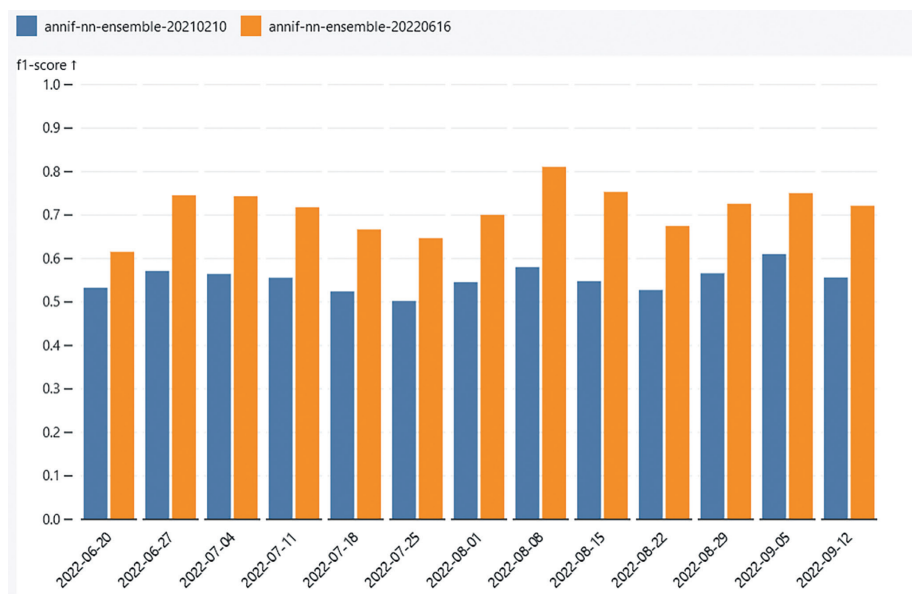
- the fact that training data is typically annotated by humans (which is also the case for AutoSE)
- the fact that knowledge organisation systems and mappings between them are usually created and maintained by humans (which applies to STW as well)
- machine-assisted subject indexing (such as machine-generated suggestions in DA-3, see above)

- and various ways of collecting intellectual feedback to approaches such as Online Learning (where a machine directly retrains itself, for example on the basis of intellectual feedback data) and Active Learning (where a machine can interactively request annotations or assessments for individual data from a human at certain points).

With respect to ways of gathering intellectual feedback, several strategies have been used in the AutoSE context. One such strategy has been conducting an intellectual review about once a year where a group of ZBW subject indexing experts assess the quality of machine-generated subjects for a sample of around 1,000 publications by assigning one of four quality levels both to each individual subject and to the sum of subjects for the publication in question. If experts found a subject missing, then they could enter that into the form as well. For this kind of review the team used an interface that was developed in project AutoIndex which allows experts to view the relevant metadata, to access the full text via a link, and to navigate in the records assigned to them (ZBW, 2022c). After every review, the team conducted an extensive debriefing where the experts could also report individual observations and perceived biases in the output of AutoSE. Over the last several reviews, this has helped to identify and to remedy systematic divergences from the desired outcome – for example, due to overrepresentation in the training data, the subjects for “theory” and “USA” wrongly appeared in the output more often than other subjects. As a temporary fix, the subject for “USA” was subsequently blocked if it was not contained in the title or the author keywords explicitly, whereas “theory” was blocked if a subject from a list with more specific subjects pertaining to economic theories compiled by the domain experts was also present in the (candidate) output. However, this approach is tedious and error-prone and such short- to medium-term solutions should be superseded by improvements in machine learning methods in the long run.

Another way of gathering feedback is comparing AutoSE output to intellectual subject indexing, where available – every time a ZBW subject indexing expert adds STW subjects to a metadata record that had already been enriched by AutoSE, the AutoSE system is notified and the F1 score is computed from the difference between the two sets of subjects. This enables the team to gather evidence that a new backend performs better than the previous one before launching it into productive operations, for example – see Figure 2 for a visualisation of an A/B test where two backends were operated

Fig. 2: Comparison of the F1 scores computed from subsequent intellectual subject indexing for two backends over time.



in parallel for a certain period of time in order to compare them with respect to this metric.

However, while the F1 score is an accepted performance indicator in classification tasks, note that this is merely a “binary” sort of feedback in the sense that it only allows to determine whether a machine-generated subject is also present in the intellectually generated set and whether an intellectually generated subject is missing from the machine-generated set but not if a machine-generated subject not chosen by the human indexer is just too general or completely incorrect, for example. Therefore, and since annual reviews yield only little feedback data due to lack of personal resources and consequently small samples, in early 2022 the team collaborated with the provider of the DA-3 platform in order to integrate a solution into DA-3 so that subject librarians could give a graded feedback. As a consequence, subject indexing experts are now able and strongly encouraged to submit quality assessments via DA-3 continuously during their everyday work, without having to change clients. As in reviews before, they can rate subjects individually and

their sum on the metadata record level. Missing subjects are computed from differences between AutoSE suggestions and the intellectual subject indexing that the experts enter into a record. The larger amount of assessment data collected this way affords the team a much better overview over AutoSE performance (as perceived by subject indexing experts) and enables them to improve their portfolio of methods in a more targeted way. Dynamically generated visualisations of this data (such as the one in Figure 2) will also be displayed via a web user interface in the future¹⁰ to increase transparency.

Future plans with respect to the implementation of a more advanced “human in the loop” relationship include exploring if this feedback data can be used for incremental learning (Online Learning). Another intriguing concept to pursue is that of Active Learning (see above). So far, automated and intellectual subject indexing represent quasi-separate lanes – machine-generated subjects are discarded as soon as human-generated ones are available (even if the latter may be inspired by the former). We would like to explore the possibilities of a more interactive mode for machines and humans to solve the task of subject indexing together that also exploits their respective strengths (better) – currently automated solutions are still designed to emulate intellectual ones as closely as possible although machines may be able to identify subtle patterns and differences where traditional rules for intellectual subject indexing are too coarse. Moreover, once the process is at least partly automated, this may also pave the way for the switch towards cataloguing and subject indexing practices that are based on entities and on formalised relationships between them, and not on (string-based) entries in a database, which in turn will facilitate the application of more advanced semantic technologies. Naturally, all these potential approaches have to be submitted to carefully designed studies as to their feasibility and actual usefulness in order to make sure that the suggested changes in information processing practices in the library are sustainable and tailored to the needs of the various users and stakeholders in a constructive way.

4. Lessons Learned

4.1. Commitment and Communication

There are many hurdles to overcome in the transfer of prototypical software results into practice – see Battistella et al. (2015) for a comprehensive

literature review about the challenges for technology transfer in general, many of which apply to our case as well.

The author of this paper has time and again heard librarians (at this institution and others) express frustration about the impression that while there may be a number of promising automation solutions from applied research activities or other institutional projects, these solutions rarely make it into productive services that they can actually use in their everyday work. One reason for such an incomplete transfer into the infrastructure of an institution may be a lack of resources, especially of personnel with suitable expertise, both for applied research and for the implementation of research results into usable software. Decision-makers in libraries often seem reluctant to commit to a larger digital transformation attempt long-term and rather prefer to label tentative in-house automation activities as “projects” in order to avoid tying down substantial amounts of resources (including *permanent* positions for highly qualified staff) for many years. This is why for AutoSE the official switch from project status to a permanent task was not just a symbolic step but essential for breaking the barrier towards going live with a first version of the service because it did signal the necessary commitment from and to everybody involved. At ZBW, the acquisition of the necessary software and computing power and the creation of an additional position for a software architect were direct consequences of that initial commitment. Nevertheless, it goes without saying that the process did require and still requires a constant renegotiation of human and material resources and dealing with shifting strategic and financial circumstances within the institution and in the world around it. With regard to staffing, experiences from the AutoSE pilot phase have shown that, at an absolute minimum, the roles that have to be filled in order to ensure permanent productive operations and a continued development of the service are those of coordination,¹¹ applied research, software development, and IT administration. The AutoSE team covers these roles with three people at the moment. Best practice also shows that at the very least the latter two types of expertise should be distributed over more than one person in order for productive operations to be fail-safe – increasing the staff within the team would obviously be an attractive option (and a luxury at most institutions) but it also helps if the technology stack of the institution as a whole is as homogeneous as possible because that facilitates the exchange of expertise with other departments, which is why inter-departmental communication and coordination is so important for the activities of AutoSE.¹²

However, commitment is not just a requirement for decision-makers. If the transfer from research all the way into productive operations is to be successful, all parties involved have to stay involved until the results are satisfactory. This also pertains to researchers – which is often problematic because the membrane between the academic world and the world of real-life use cases still seems to act as an obstacle to fruitful cooperation. Research processes in computer science can be categorised very roughly into two types: fundamental research yields theoretical findings which stay true no matter how they are applied, while applied research yields prototypical software which then has to be transformed into productive services. The execution of that latter step heavily depends on the application context and is not a one-way street but requires many cycles to adjust the outcome to the use case. These subsequent stages do not only involve additional scientific tests, but also the development and testing of software for productive operations, doing usability tests with the target users in order to find further issues and to ensure acceptance, and so on.¹³ Typically, these are the stages where the transfer process is prone to seizing up and the people involved get frustrated.

One cause for this could be that in the past, the latter stages have been underestimated by both decision-makers and researchers and as a consequence, success in this domain is not rewarded with prestige the same way as academic success. Sadly, there is currently no real incentive for researchers to stay involved beyond the prototype stage – on the contrary, the pressure caused by academic key performance indicators (“number of publications”, “amount of third-party funding acquired”, etc.) is so prohibitively intense that it actively keeps them from doing so. AutoSE has had the fortune of having been assigned a full PhD position within the AutoSE team so that that staff member could focus both on their research and on the application context and thus create as much synergy as possible – however, when collaborating with other researchers outside of the team these conflicting priorities are still a challenge that can sometimes get in the way.¹⁴ A (partial) solution would be for policy-makers to create key performance indicators that measure and reward successful research transfer activities (both for institutions and for individuals involved in such activities) and to attribute the same significance to them as to the other indicators.

Shifting more attention towards the relevance and complexities of the latter stages would hopefully a) cause decision-makers to provide sufficient human resources to fill all the essential roles (see above), and b) increase the extent to

which the skills necessary for those latter stages are taught during the education of prospective applied researchers. This includes both the technical skills needed to participate in large-scale software development projects (cooperative programming, testing, deployment, and so on) and the soft skills needed for project management and for communicating with stakeholders, so that researchers are endowed with the necessary toolkit to find out where the practical challenges lie and to help solve them. In essence, all parties – applied researchers, software engineers, stakeholders – have to continuously communicate requirements and issues back and forth in order to effect a successful transfer of research results into a productive service that is suitable and equipped for permanent operation.

4.2. Conclusion

Experiences from the pilot phase to date have shown the following: As of yet, there are no shelf-ready open source automation systems for subject indexing – existing software has to be adapted and maintained continuously which requires various forms of expertise. The step of leaving the project format behind is worth the effort – the search for automation solutions for subject indexing and other related processes is a permanent task that will stay with libraries for many years to come. Accordingly, productive operations in line with this task have to be based on a thoroughly established long-term concept and to be accompanied by adequate resources (personnel, software, computing power).

We have found the fact that applied research and software development for AutoSE is done *within* the library part of ZBW (and not in a separate research or IT development department) greatly beneficial because it allows a close collaboration and communication with subject librarians. It is essential to include subject indexing experts as stakeholders in the process – both for their expertise in the areas of information and knowledge organisation and to increase acceptance since transparency helps to dissipate reservations and to establish a basic trust in the technology and especially in the ways the team is going to use it. The implementation of methods from Artificial Intelligence can assist libraries in their continued mission to prepare and provide information resources while remodelling their information processing practices in a novel way. The concept of *human in the loop* offers a possible approach

for retaining intellectual subject indexing expertise while combining it with machine-learning-based methods and thus transferring it into a form that is more adapted to the potential of state-of-the-art technology available today.

References

- Battistella, C., De Toni, A., & Pillon, R. (2015). Inter-organisational technology/knowledge transfer: a framework from critical literature review. *The Journal of Technology Transfer*, 41, 1195–1234. <https://doi.org/10.1007/s10961-015-9418-7>
- Bus factor. (2023, July 5). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Bus_factor&oldid=1154218987
- Dong, T., & Suominen, O. (2022). *Omikuji* [Computer Software]. <https://github.com/tomtung/omikuji>
- Eurospider Information Technology AG. (2023, January 17). *Subject indexing using the DA*. <https://www.eurospider.com/en/relevancy-product/digital-assistant-da-3>
- F-score. (2023, January 17). In *Wikipedia*. <https://en.wikipedia.org/wiki/F-score>
- Facebook Inc. (2022). *fastText* [Computer Software]. <https://fasttext.cc/>
- German National Library. (2022, October). *Automatic Cataloguing System*. https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/KI/ki_node.html
- Grote, C. (2022, May 11). *German National Library launched its “cataloguing machine”* [Online forum post]. Annif Users. <https://groups.google.com/g/annif-users/c/KVQB-hvLrbA/m/19RwM9EPBgAJ>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). *Bag of tricks for efficient text classification*. arXiv preprint. <https://doi.org/10.48550/arXiv.1607.01759>
- Kasprzik, A. (2022, July 26–29). *Get everybody on board and get going – the automation of subject indexing at ZBW* [Conference presentation] 87th IFLA World Library and Information Congress (WLIC), Satellite Meeting: Information Technology – New Horizons in Artificial Intelligence in Libraries, Dublin, Ireland. <https://repository.ifla.org/handle/123456789/2047>
- Monarch, R.M., & Manning, C.D. (2021). *Human-in-the-loop machine learning – active learning and annotation for human-centered AI*. Manning Publications. <https://livebook.manning.com/book/human-in-the-loop-machine-learning/>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Suominen, O., Inkinen J., Virolainen, T., Fürneisen, M., Kinoshita, B.P., Veldhoen, S., Sjöberg, M., Umstein, P., Neatherway, R., & Lehtinen, M. (2023a). *Annif* [Computer Software]. <https://doi.org/10.5281/zenodo.5654173>

Suominen, O., Inkinen J., Virolainen, T., Fürneisen, M., Kinoshita, B.P., Veldhoen, S., Sjöberg, M., Umstein, P., Neatherway, R., & Lehtinen, M. (2023b). *nn-ensemble* [Computer Software].

Toepfer, M., & Seifert, C. (2018a). Fusion architectures for automatic subject indexing under concept drift. *International Journal on Digital Libraries*, 21, 169–189. <https://doi.org/10.1007/s00799-018-0240-3>

Toepfer, M., & Seifert, C. (2018b). Content-based quality estimation for automatic subject indexing of short texts under precision and recall constraints. In E. Méndez, F. Crestani, C. Ribeiro, G. David, & J. Correia Lopes (Eds.), *Lecture Notes in Computer Science: Vol. 11057. Digital Libraries for Open Knowledge* (pp. 3–15). Springer. https://doi.org/10.1007/978-3-030-00066-0_1

Transformer (machine learning model). (2023, January 17). In *Wikipedia*. [https://en.wikipedia.org/wiki/Transformer_\(machine_learning_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))

ZBW – Leibniz Information Centre for Economics. (2022a). *stwfsapy* [Computer Software]. <https://github.com/zbw/stwfsapy>

ZBW – Leibniz Information Centre for Economics. (2022b). *qualle* [Computer Software]. <https://github.com/zbw/qualle>

ZBW – Leibniz Information Centre for Economics. (2022c). *releasetool* [Computer Software]. <https://github.com/zbw/releasetool>

ZBW – Leibniz Information Centre for Economics. (2023, January 17). *STW Thesaurus for Economics*. <https://zbw.eu/stw/version/latest/about.en.html>

Notes

¹ The author of this paper.

² See <https://github.com/NatLibFi/Annif/wiki/Backend%3A-Omikuji> for its integration into Annif.

³ See <https://github.com/NatLibFi/Annif/wiki/Backend%3A-fastText> for its integration into Annif.

⁴ See <https://github.com/NatLibFi/Annif/wiki/Backend%3A-STWFSA> for its integration into Annif.

⁵ Experiments in the AutoSE context have shown that author keywords improve the F1 score (“F-score”, 2023) to 0.55 on average as opposed to 0.47 when only using titles.

⁶ For example, there had been a regular exchange of ideas prior to the German National Library launching their own Annif-based “cataloguing machine” and associated AI project, see Grote (2022) and German National Library (2022).

⁷ The number for intellectual subject indexing at ZBW is around 30,000 records per year.

⁸ ~147,000 in 2021; ~500,000 in 2020.

⁹ For a short description see Eurospider Information Technology AG (2023).

¹⁰ (January 2023) A prototypical version exists – the next step is to launch the UI internally.

¹¹ In Battistella et al. (2015) this role is called the “intermediary”: The intermediary “acts as a third party agent assuming the role of facilitation/mediation between the parties in order to facilitate the relational context and with the aim of supporting the development of the process in its criticalities, addressing enabling or constraining factors.”

¹² The goal is to increase the so-called “bus factor” (2023) of the team which is “the minimum number of team members that have to suddenly disappear from a project before the project stalls due to lack of knowledgeable or competent personnel”.

¹³ “Technology transfer is a bilateral process between sender and receiver: there is a ‘process of feed-back’ from the sender to the receiver, which allows interested parties to obtain more information (knowledge) on the use of technology transferred” (Battistella et al., 2015) – the same is true for applied research results.

¹⁴ In terms of the challenges compiled by Battistella et al. (2015) this could also be interpreted as an issue of “cultural distance”, i.e., the transfer gets harder if there is a certain lack of a shared vision or a common goal.