SN Computer Science (2025) 6:813
https://doi.org/10.1007/s42979-025-04348-9

ORIGINAL RESEARCH

AutoTag & *TagMap*: LLM-Powered Moodle Plugins for Pedagogical Alignment Checks

Christian Grévisse[1] · Claude Braun[2] · José Batista da Costa[3]

Received: 24 July 2025 / Accepted: 23 August 2025

Abstract
Pedagogical alignment denotes the coordination between learning outcomes, teaching and learning activities and assessment tasks. For any learning outcome of a course, teachers should design corresponding learning activities and consequently assess its achievement. Fair and valid assessments require a proper pedagogical alignment. At a low-level, this can be verified by tagging learning material, e.g., lecture notes or presentations, and assessment items, e.g., quiz questions, and checking a homogeneous coverage. Learning Management Systems (LMS) such as Moodle enable teachers to tag both resources and questions. To ease the tagging, named entity recognition technologies can be used. With the advent of Large Language Models (LLMs) such as GPT-4, this task has seen a new momentum. In this paper, we present *AutoTag* and *TagMap*, two Moodle plugins which will help teachers to check the pedagogical alignment of their online courses. The first plugin, *AutoTag*, leverages GPT-4 to provide automatic resource tagging support to teachers. An existing plugin for question generation also implements this logic. The second plugin, *TagMap*, independent of the other, visualizes concept coverage in both resources and questions to help teachers in verifying the pedagogical alignment and identifying possible shortcomings. We exemplify the usage of both plugins in a nephrology and urology course.

Keywords Pedagogical alignment · Moodle · Large Language Models

Introduction

In teaching and learning, *pedagogical alignment* is a cornerstone concept. It denotes the coordination between the intended learning outcomes, the teaching and learning activities and the assessment tasks [2, 3]. For any given learning outcome, teachers need to create learning activities that can help students achieve this outcome as well as assessment tasks to check whether it actually got achieved. For instance, in a medical education course where students should learn to distinguish between normal and pathological renal and urinary conditions, explaining to students the concept of *hematuria* and assessing their knowledge with a clinical case involving this condition, can be considered as a good alignment. Caution should be raised when too many concepts that heavily occur in learning material are not assessed, indicating a gap between instruction and assessment. Even more problematic is when the assessment does not truly evaluate what was taught. This could occur if concepts covered by questions were not subject of the instruction and cannot be assumed as prior knowledge.

An example is shown in Fig. 1. Learning material from a teacher, in this case a presentation, covers different anomalies of urine composition and color. A subset of the concepts associated to these slides comprises *hematuria* and *food impact*. A related assessment item also covers *hematuria*, but could also be associated to obesity. The slides

Claude Braun and José Batista da Costa contributed equally to this work.

✉ Christian Grévisse
christian.grevisse@uni.lu

Claude Braun
claude.braun@hopitauxschuman.lu

José Batista da Costa
jose.batistadacosta@hopitauxschuman.lu

1 Department of Life Sciences and Medicine, University of Luxembourg, 2, place de l'Université, 4365 Esch-sur-Alzette, Luxembourg

2 Department of Nephrology, Hôpitaux Robert Schuman, 9, rue Edward Steichen, 2540 Luxembourg, Luxembourg

3 Department of Urology, Hôpitaux Robert Schuman, 9, rue Edward Steichen, 2540 Luxembourg, Luxembourg

Published online: 10 September 2025

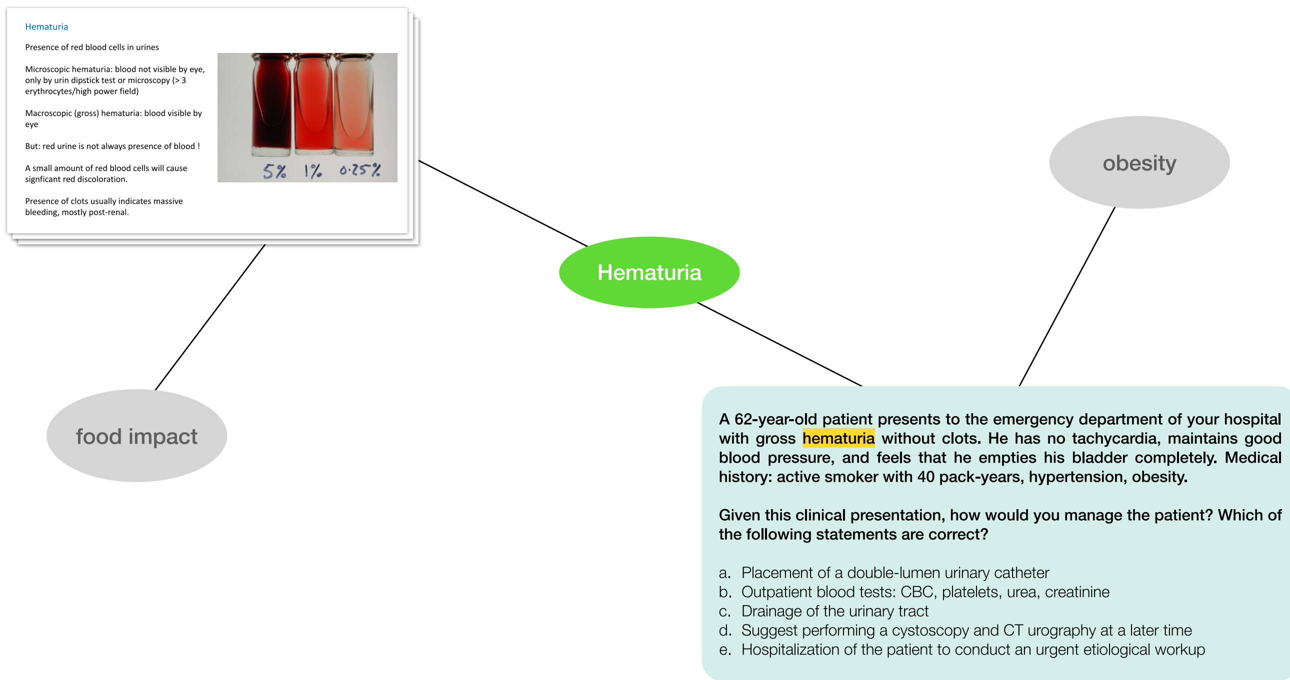


Fig. 1 Example of learning material, an assessment item and related concepts

and the question have the *hematuria* concept in common, so the concept is both covered and assessed. The *food impact* concept is not associated to an assessment item, most likely because urine color anomalies due to, e.g., eating beets, is considered benign and temporary. Nonetheless, this could give teachers an insight into the opportunities for further assessment items. On the other hand, *obesity* is not covered by the learning material of the course, probably because this is not a main topic of a nephrology/urology course and would rather be covered in a nutrition course. Again, teachers could benefit from this realization and address this topic in class.

Contemporary Learning Management Systems (LMS) typically enable teachers to associate concepts to learning material and assessment items through *tags*. The popular Moodle[1] LMS, at the time of writing presenting over 150,000 installations, introduced tagging support for *Activities* and *Resources* (e.g., files) in its version 3.1 released in May 2016. Tagging of quiz questions is also supported. However, tagging resources or questions from scratch requires significant effort from teachers.

This task can be eased by taking advantage of *Named Entity Recognition* (NER) [5], sometimes called *Core Concept Identification* (CCI) [10] or *Automatic Term Extraction* (ATE) [1]. This field traditionally used Natural Language Processing (NLP) technologies, further augmented by Semantic Web tools, such as *DBpedia Spotlight*.[2] With the advent of *Large Language Models* (LLMs) such as `GPT`[3]`-4`, NER has seen a shift towards *in-context learning*.

In this paper, we present *AutoTag* and *TagMap*, two Moodle plugins that in combination allow to check the pedagogical alignment of an online course by identifying concepts as well as comparing their coverage among contained resources and questions. The *AutoTag* plugin leverages GPT-4 to provide automatic resource tagging support to teachers. An existing plugin for question generation also implements this logic. The *TagMap* plugin then visualizes the coverage of concepts in both resources and questions to help teachers in verifying the pedagogical alignment. This latter plugin is independent and could be used without the LLM-based automatic tagging support. The main contribution of this work consists of two Moodle plugins available to the community as well as a case study that shows their application in a undergraduate nephrology-urology course.

[1] https://moodle.org.

[2] https://www.dbpedia-spotlight.org.

[3] Generative Pre-trained Transformer.

Related Work

LMS Tagging Support

Before the introduction of tagging support for course modules (i.e., activities and resources) in Moodle 3.1, Elmadani et al. developed a Moodle plugin for concept tagging of course modules [4]. The authors emphasized the benefits of concept tagging as this metadata can be used to track student progress and to suggest further course modules.

Mazza et al. also focussed on the importance of tags for students [11]. If a course contains a high number of resources, it will be difficult for students to find the right resource for a certain concept. Furthermore, if the title is not meaningful either (e.g., "Lesson 1"), students will need to open every file to search for the right resource.

García-González et al. developed *LODLearning*, a prototype for the Sakai LMS[4] that links latent content into existing courses [6]. Upon file upload, an entity recognition algorithm within *Apache Stanbol*[5] is executed and the resulting entities are shown to the teacher. The tool then leverages the Linked Open Data Cloud to link the recognized entities to Wikipedia information.

Grévisse et al. extracted concept annotations from file metadata to add them as tags in Moodle [9].

LLM-Based NER

Although LLMs can benefit from fine-tuning, where the internal weights are adapted according to a training set, the more common approach lies in in-context learning. Zero-shot learning relies on a prompt that only contains instructions. Few-shot learning integrates examples in the prompt, which can be useful depending on the task at hand.

Banerjee et al. compare ATE results for a few-shot approach with GPT-3.5-Turbo against a fine-tuned BERT-based model [1]. GPT-3.5-Turbo outperformed the BERT-based model. It inherently overcomes the costly supervised approach of a BERT-based model, as the annotation effort is significant. There can also be domain-specific variations, but GPT-3.5-Turbo was particularly good in the heart failure domain. Both models may suffer from a high recall and a significant number of false positives (e.g., returning non-domain-specific terms).

García-Barragán et al. used GPT for medical entity recognition in Spanish electronic health records (EHR) related to breast cancer [5]. They employed in-context learning with external knowledge (i.e., providing a vocabulary of entities to use). In comparison to BERT, GPT showed a comparable accuracy. Zero-shot learning with GPT-4 and external knowledge resulted in the highest precision, 5-shot with external knowledge achieved the highest F-score compared to GPT-3.5. Again, the authors emphasize the advantage of GPT not requiring the costly pre-annotation compared to BERT. Similar results were also achieved with GPT for phenotype concept recognition [8].

Moore et al. used GPT-4 for automated generation and tagging of Knowledege Components (KCs) for Multiple-Choice Questions (MCQs) in the chemistry and e-learning domains [12]. KCs can be understood as skills, competencies or concepts. The creation of KCs, as any tagging activity, is time-intensive and requires domain expertise. The authors compared the generated KCs to human-assigned ones, and upon differences, domain experts leaned towards the generated ones. The authors propose a *human-in-the-loop* approach, in which a domain expert reviews and either confirms or modifies generated KCs. However, the initial generation "*significantly reduces the time and effort required compared to starting from scratch*".

To the best of our knowledge, this work is the first to integrate LLM-based tagging support for learning material and questions in the Moodle LMS as well as the first to visualize the coverage of concepts across resources and assessment items.

Moodle Plugins

In this section, we will present the two Moodle plugins we developed, *AutoTag* and *TagMap*. Like any Moodle plugin, they are written in PHP and can be used from Moodle 4.1 onwards.

AutoTag

The *AutoTag*[6] plugin helps teachers in adding tags with reduced effort. Apart from teachers being able to check, based on these tags, the pedagogical alignment using the *TagMap* plugin, students will also benefit from the tagging in order to discover learning material related to a given concept through, e.g., the standard Tag block integrated in Moodle.

Why is this tagging support for teachers needed? The current Moodle instance at the University of Luxembourg contains 3750 visible courses, containing a total of 99,732 course modules (i.e., activities or resources). On average, there are 27 course modules per course, with a median value of 14. This adds to the challenge that students may face to find the right resource for the right concept. Almost

[4] https://www.sakailms.org.

[5] https://stanbol.apache.org.

[6] https://github.com/cgrevisse/moodle-local_autotag.

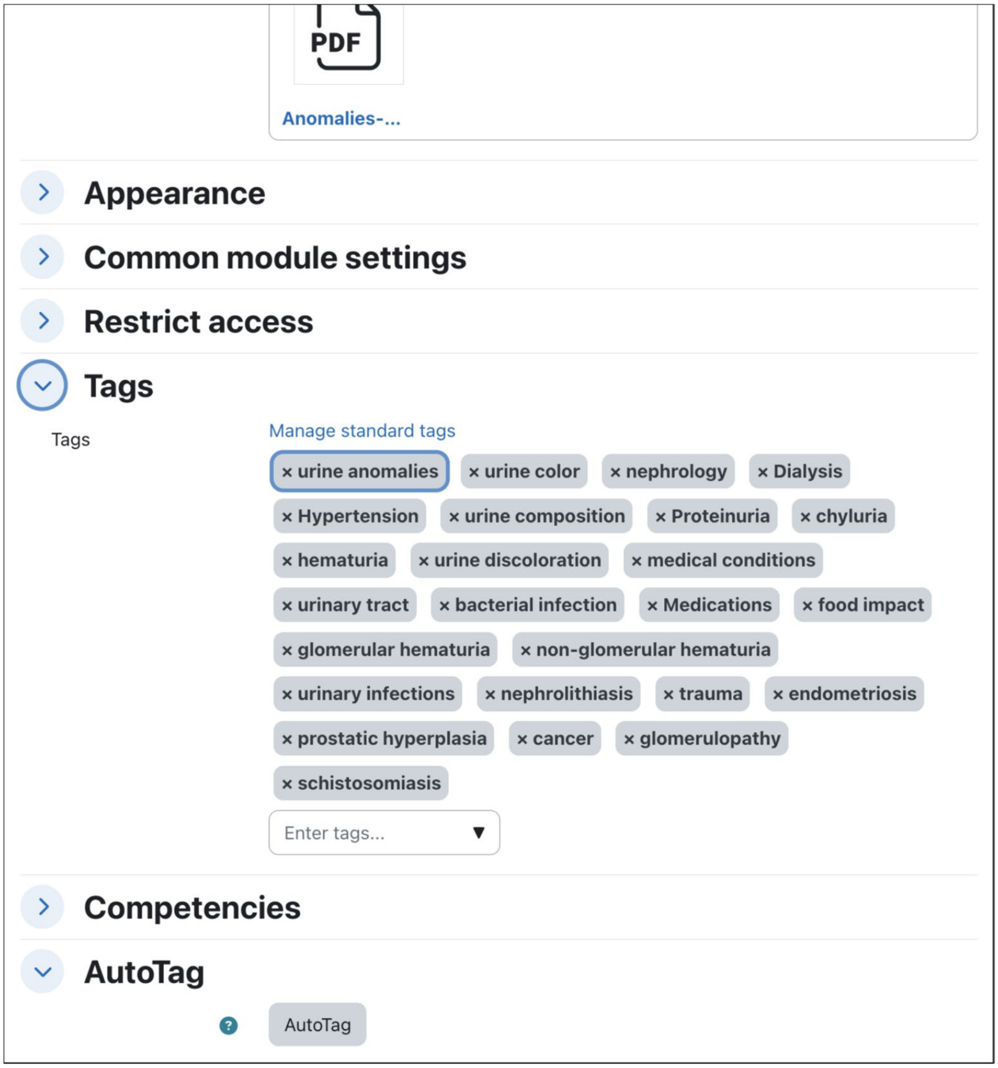


Fig. 2 AutoTag button and resulting tags

Fig. 3 System prompt for the AutoTag plugin

> **System prompt**
>
> `You are a tagging assistant. Your task is to extract a list of the most important tags for the given content. All tags shall be given in English.`

two thirds of course modules are resources (i.e., files), the majority being PDF files. Furthermore, the platform comprises 2648 tags, mostly used for user profiles and quiz questions. There are 149,081 quiz questions, of which 1837 (1.2%) are tagged. Among these, the number of tags per question shows a third quartile of 1. Only 48 course modules (0.04%) are tagged, half of which only contain a single tag. This data shows a clear underusage of the tagging feature, likely due to the additional effort required by teachers.

The plugin can be employed by users with the necessary capability to add file resources to a Moodle course. Inside the settings of a resource, an *AutoTag* button allows

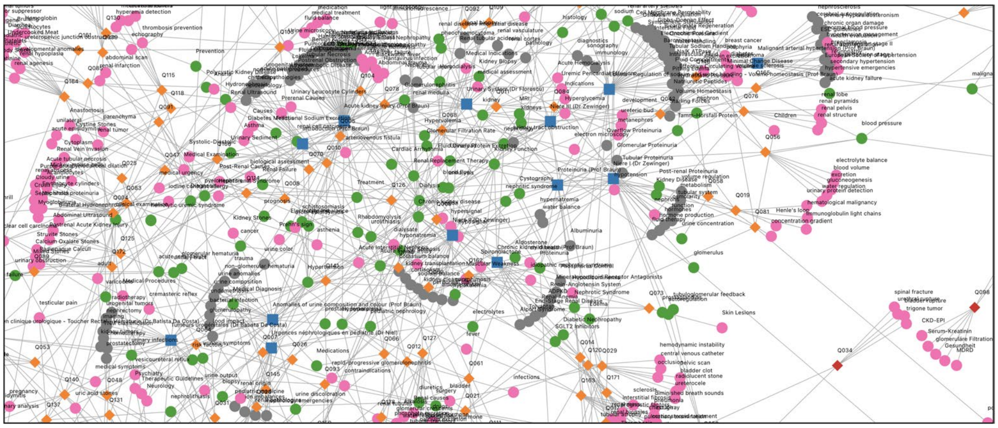


Fig. 4 Graph visualization of pedagogical alignment

users to start a background task that will send the resource to the *Responses API*[7] of OpenAI (Fig. 2). The system prompt is shown in Fig. 3. English is used as a common denominator here, as courses could contain resources in multiple languages.

The user prompt is essentially the file content, sent in Base64 format. At the time of writing, only PDF files are supported by the file input of the Responses API. Structured output is guaranteed through the usage of a JSON schema, resulting in an object as given in Listing 1. The *temperature*, hyperparameter indicating the randomness of the output, is set to 0, ensuring a more deterministic result. As shown in Fig. 2, the returned tags are then added to the tag input, such that teachers can review, add, modify or delete them. The plugin settings let site administrators add an OpenAI API Key.

A similar logic was introduced in the existing *Generative AI Question Bank* plugin.[8] This plugin uses GPT-4 for question generation based on course resources and was evaluated in the computer science and medical domains [7]. Questions in the question bank can now be tagged individually or in a bulk action, provided that the user has the capability to tag (all) questions. The same system prompt as in Fig. 3 is used. The user prompt contains the question text and, in case of a multiple-choice question, the answer options. Again, structured output with JSON schema and a 0 temperature are used.

Through the new *AutoTag* plugin and the addition in the existing question generation plugin, teachers can now benefit from assistance for tagging learning material and questions. The addition of tags, apart from the already mentioned benefits for students, will eventually help teachers to determine

Listing 1 Structured output example

```
{
    "tags": ["Proteinuria", "Nephrology", "Kidney Disease
        ", "Glomerular Proteinuria", "Tubular Proteinuria"]
}
```

7 https://platform.openai.com/docs/guides/text.

8 https://moodle.org/plugins/qbank_genai.

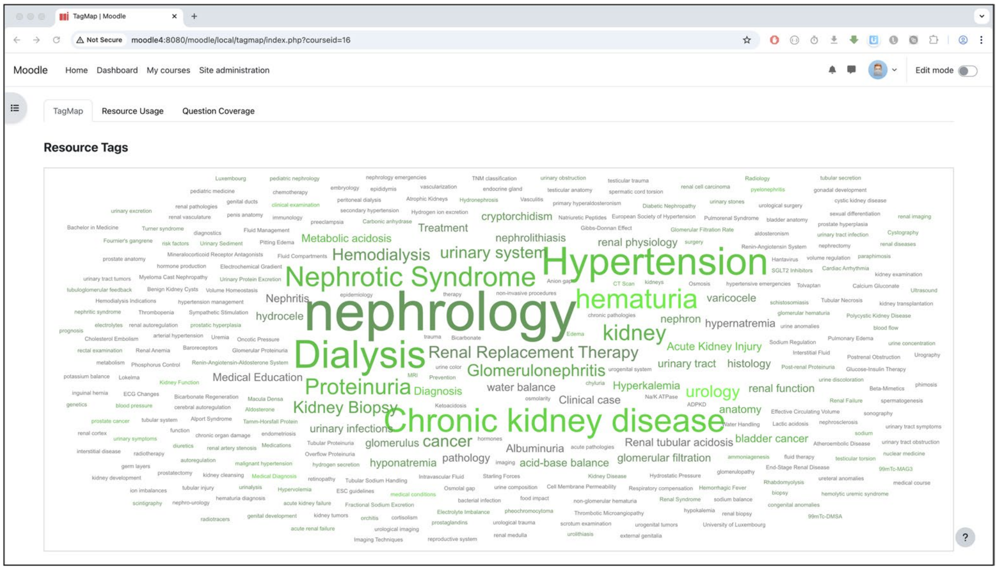


Fig. 5 Word cloud representing the resource tags

the pedagogical alignment between instruction and assessment, fostered by the *TagMap* plugin.

In both plugins, we used the `gpt-4o`[9] model. Note that the employed prompts leverage zero-shot learning. While this approach bears a higher risk of high recall and false positives [1], the initial set of generated tags constitutes a support to teachers, as tagging is a time-intensive task [12]. Furthermore, the usage of LLMs avoids costly annotation as required by traditional supervised learning approaches [1, 5].

TagMap

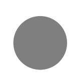
The *TagMap*[10] plugin visualizes the pedagogical alignment between instruction (learning material) and assessment (quiz questions). Our working hypothesis is that the alignment can be determined by comparing the concepts covered in learning material and quiz questions, enabling teachers to identify those concepts that are either taught but not assessed or assessed but not taught. The plugin can benefit from the tagging support of the *AutoTag* and *Generative AI Question Bank* plugins, but it does not need to. If teachers have already tagged resources and questions, either manually or through different means, the *TagMap* plugin can still be used as a standalone plugin and without any interaction with LLMs. The plugin is teacher-centric, requiring the capabilities to view all questions in the question bank and view hidden course modules. The `D3.js`[11] library for data visualization is used.

In a first approach, we created a graph visualization (Fig. 4). Resources, tags and questions represent the nodes, connected by edges if a coverage relation between a tag on one hand and a resource or question on the other hand exist.

The different types of nodes can be explained as follows:

- Resources are represented as blue squares.
- Tags are of circular shape:

  – Pink if not covered by any learning material;



  – grey if not covered by any quiz questions;

  – green if covered by both learning material and questions.

- Questions are of diamond shape:

[9] More precisely: `gpt-4o-2024-08-06`.

[10] https://github.com/cgrevisse/moodle-local_tagmap.

[11] https://d3js.org.

– Red if no associated tag is covered by any learning material;

– orange for partial coverage;

– green if all tags are covered by learning material.

The map is zoomable and draggable but it is not very intuitive due to the high number of nodes and overlapping edges. In addition, rendering times are rather important.

In a second approach, we used word clouds by leveraging the `d3-cloud` plugin.[12] We provide different views:

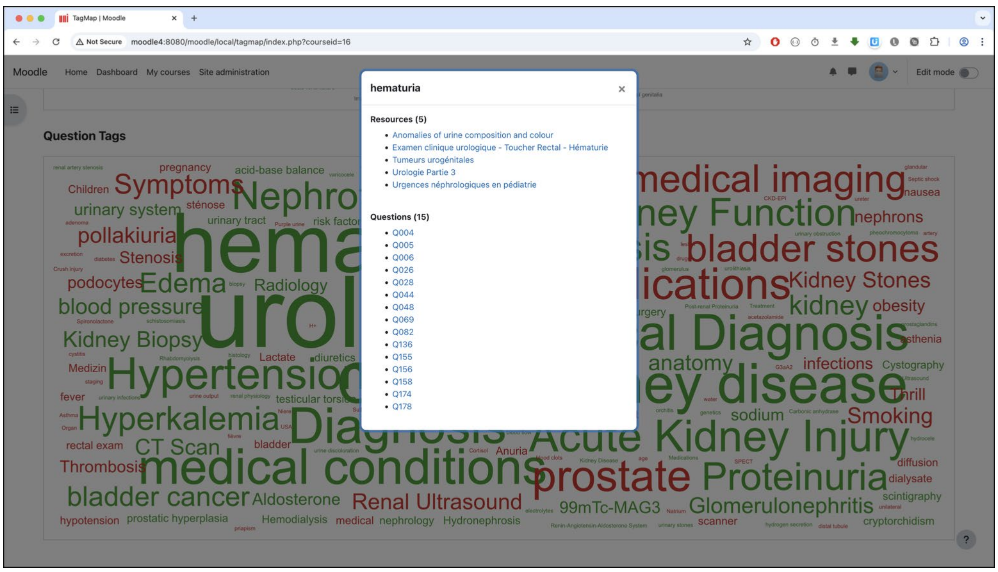


Fig. 6 Word cloud representing the question tags

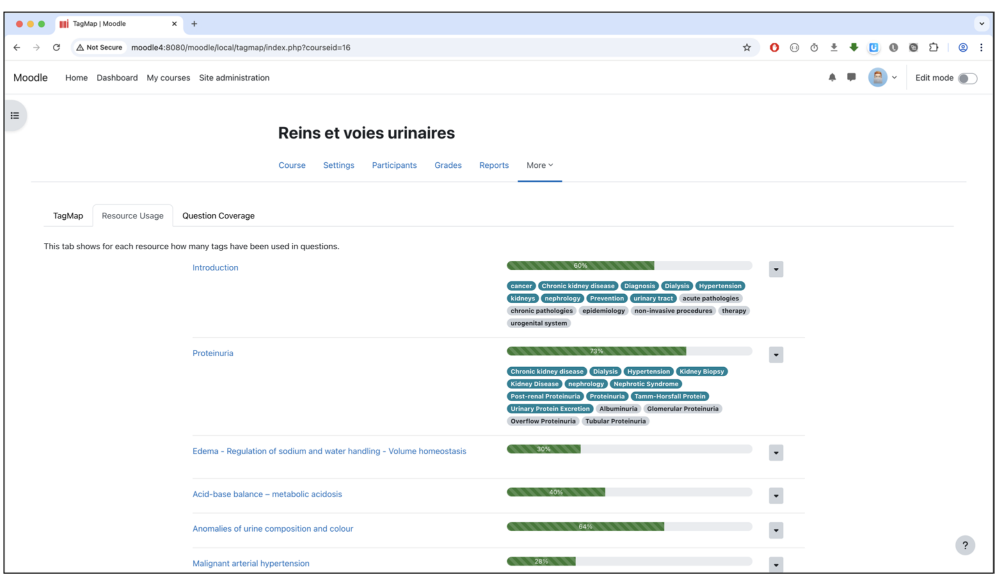


Fig. 7 Resource Usage: list of resources and usage of associated tags in questions

[12] https://github.com/jasondavies/d3-cloud.

- **Resource Tags:** In this word cloud (Fig. 5), the size of a tag represents its frequency inside the resources of a course, whereas the color indicates the coverage within questions. Here, we interpolate between grey (not used at all) and fair green (used by a maximum of questions per tag).
- **Question Tags:** In this word cloud (Fig. 6), the size of a tag represents its frequency among questions, whereas the color indicates the coverage inside learning material. If the concept is tagged in at least one resource, it is shown green. Red concepts are not covered at all.

For both word clouds, clicking on tags will open a modal showing all related resources and questions with links to the files respectively question previews. This will ease the navigation for teachers, e.g., to check whether the concept covered in a question but not in learning material is highly relevant and should thus be addressed in class.

The word clouds are rather tag-centric, so we provided two further views for teachers:

- **Resource Usage:** This view (Fig. 7) shows for each resource the corresponding tags and whether they are used in questions. This view indicates teachers to what extent the content of the learning material is assessed.
- **Question Coverage:** Inversely, this view (Fig. 8) shows to what extent tags associated to each question are covered by learning material. Teachers can thus check whether questions are appropriate with respect to the instruction.

In both views, clicking on tags will lead to the same modal as shown in Fig. 6.

Case Study

As seen in the previous figures, we tested the plugins in a nephrology-urology course in the undergraduate medical program at the University of Luxembourg. Previous work already highlighted the good performance of GPT for named entity recognition in the medical domain [1, 5]. The course is taught in English, French and German, with learning material and quiz questions in these three languages. While this might seem challenging, GPT already showcased its capacity for medical entity recognition in Spanish [5]. By asking the model in the system prompt to return all tags in English (Fig. 3), we try to homogenize the situation.

There were 19 resources (8 in English, 8 in French and 3 in German). The question bank contained 178 questions. In total, 1043 tags were identified, with a median of 18 tags ($\mu = 20.1 \pm 8.4$) per resource and 7 ($\mu = 7.1 \pm 2.2$) per question. There is a median of 0 resources but 1 question per tag. 12% of the tags were covered and asked, 16% were covered but not asked and 72% were asked but not covered. Assuming these were the final tags, this would indicate the need to better cover certain concepts in class. However, a variation in granularity could also falsify the picture. For instance, *Proteinuria* is a broader concept for *Glomerular Proteinuria* and *Tubular Proteinuria*. Clustering concepts according to hierarchical relations could help here.

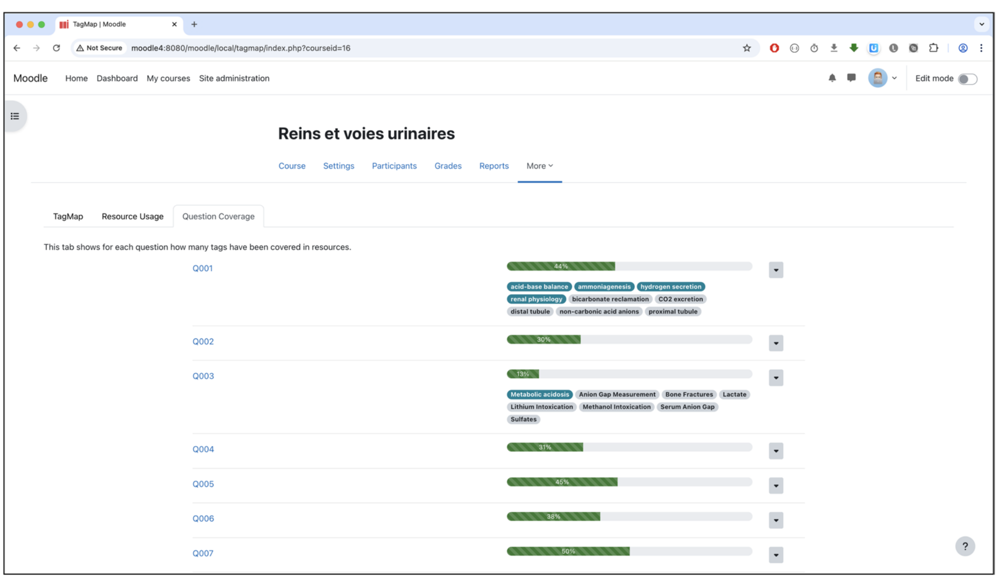


Fig. 8 Question Coverage: list of questions and coverage of associated tags in learning material

Two domain experts, a nephrologist and a urologist checked the corpus of resources and questions for appropriate tagging. More concretely, they were asked to remove concepts that they did not deem as a core concept of the given resource or question and to add any potentially missing ones. For resources, the average number of deleted tags ($\mu = 10.2 \pm 7.4$) was significantly higher than the number of tags that were added by the teacher ($\mu = 0.2 \pm 0.5$). This translates into a moderate average *precision* of 0.53 and a high average *recall* of 0.99. The situation is similar for questions, where an average of 4.0 ($\pm$ 2.3) tags got excluded and 0.2 ($\pm$ 0.4) got added by the first teacher. Here, the average precision dropped to 0.45 and the average recall to 0.93. A subset of 35 questions also got evaluated by the second teacher, with an average of 3.3 ($\pm$ 2.0) tags deleted and 0.8 ($\pm$ 0.8) added, resulting in a comparable precision (0.47) but a lower recall (0.74). For these questions, the root mean square error (RMSE) between both raters' precision and recall values was 0.35 respectively 0.38, indicating a certain disagreement. As only two evaluators with a small subset of overlapping ratings can be reported, the interpretability and representativeness of these results are limited.

The low to moderate precision indicates that a lot of false positives were included by the LLM in this zero-shot learning approach. This could be improved by the pipeline proposed in [13], where both the original content and a generated summary were annotated using DBpedia Spotlight. This approach resulted in a significantly higher F_1 score compared to zero-shot learning. Nevertheless, the effort for deleting inappropriate tags is lower than for teachers to start tagging from scratch or to add missing tags. Fortunately, the high recall indicates that this was not often required. The phenomenon of a high recall and a significant number of false positives for GPT-based ATE were reported in similar work [1]. This also joins the human-in-the-loop system proposed in [12], in which domain experts gain time through the initial generation and only review and confirm the appropriateness of labels.

A question that arose during the evaluation is whether only *core* concepts should be considered for tagging. On one hand, using all recognized concepts as tags can lead to a cluttered situation as depicted in Fig. 4. A too fine-grained tagging can lead to a wrong impression with respect to the concept coverage. On the other hand, any concept appearing in an assessment item should be known and have been covered, otherwise the fairness of the item could be compromised. Not tagging them could lead to oversights. This trade-off needs to be taken into consideration when tagging resources and assessment items.

What insights can teachers gain from the *TagMap*? In the resource tag cloud (Fig. 5), generic terms like *nephrology* can be sorted out, as they are not providing further evidence in the context of this course. However, it is good to see that *hematuria* appears often in resources *and* is highly covered in questions. Smaller tags, i.e., concepts not often covered in resources, are mostly grey, so not often used in questions. This shows that less importance is given to them. Ideally, bigger tags should tend towards green, whereas smaller ones should remain grey.

For the question tag cloud (Fig. 6), generic terms could again be filtered out. It is positive that concepts like *proteinuria* or *hyperkalemia*, often coming up in questions, were covered by material. For non-covered concepts, it has to be seen whether they are covered in other courses (e.g., *prostate* is likely to have been mentioned in an anatomy course) or whether there is a real need to cover them more deeply in this course. This constitutes a real help for teachers to avoid unfair assessment items. Here, ideally, all tags should tend towards green.

Discussion

The combination of the *AutoTag* and *TagMap* plugins can indeed help teachers in quality assessment and course redesign. For instance, the *TagMap* visualization revealed that

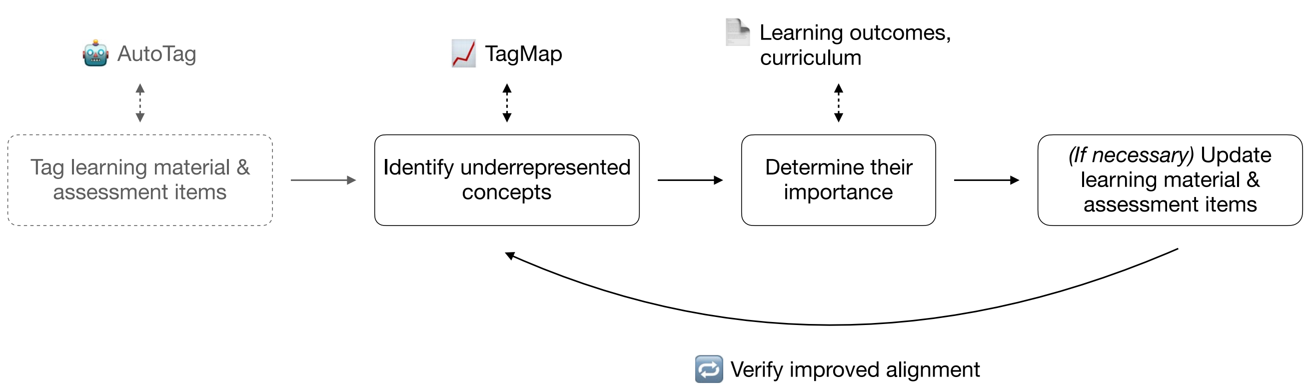


Fig. 9 Redesign workflow

obesity was assessed but not taught. This insight could prompt instructors to either remove the question or add relevant instructional material, if the concept has not already been covered in previous courses.

The redesign workflow is shown in Fig. 9. First, *AutoTag* is used to tag learning material and assessment items. This step is optional, as the tagging could also have been carried out by the teacher beforehand. *AutoTag* reduces the tagging effort by providing an initial set of identified concepts. In the next step, the different visualizations in *TagMap* help to identify underrepresented concepts, i.e., concepts assessed but not taught, or taught but not assessed. Then, the teacher needs to determine whether any action is necessary by determining the importance of these concepts, validating against the learning outcomes of this course and the overall curriculum. This action is particularly important for courses where many teachers participate and where the course coordinator needs to check the overall consistency and fairness. If any action is deemed necessary, the teacher can then update the learning material or assessment items accordingly, i.e., by adding further resources or changing questions if concepts were assessed but not taught, or modifying resources or adding questions if concepts were taught but not assessed. These steps can be repeated to verify the improved alignment.

The two plugins may benefit different stakeholders. Teachers can save time through automated tagging support, identify instructional gaps and improve fairness and validity of assessments. Students can benefit from the tagging to easier navigate and identify relevant resources via concept-based search. They also benefit from the improved pedagogical alignment through better fairness and validity of assessments. Instructional and curriculum designers can use the tag data to evaluate course coherence and be guided in course redesign. Finally, institutions can use the tag-based analytics for quality assurance and reporting, accreditation processes or curriculum reviews.

Conclusions and Future Work

In this paper, we presented *AutoTag* and *TagMap*, two Moodle plugins which will help teachers to check the pedagogical alignment of their online courses. Leveraging the power of LLMs, teachers can now benefit from an automatic resource and question tagging support. Concept coverage visualization in both resources and questions helps teachers in verifying the pedagogical alignment and identifying possible shortcomings. We exemplified the usage of both plugins in a nephrology and urology course. Students will also indirectly benefit from the tagging support for teachers, as this metadata facilitates the retrieval of resources of interest among a potentially high number of course modules. While the zero-shot approach might still result in a high number of tags, teachers are provided assistance here.

There are several possibilities for future work. Other taggable elements within a Moodle course, such as glossary entries or wiki pages, could be included for automatic tagging support. This addition in the *AutoTag* plugin could be achieved by passing the content of such glossary or wiki resources to an LLM, in the same way quiz questions are passed. A curriculum-wide dashboard could also be interesting for determining the pedagogical alignment across a series of courses. This holistic picture could show the dependencies of concepts introduced in this series of courses, and allow to understand whether a concept was already covered in a previous course and can thus be considered as prior knowledge. Gaps or redundancies in the curriculum could also be identified. Applying our approach in domains other than medicine is another potential direction. While the medical domain comprises a vast set of well-defined and delimited concepts, which is also reflected in the significant number of controlled vocabularies such as MeSH,[13] courses in the humanities might face a different situation. Trying different LLMs, such as the recent GPT-5, could also be interesting. Finally, we will submit the two plugins for inclusion in the Moodle Plugins directory.

Funding No funding received.

Data Availability No dataset was produced. The Moodle plugins are available on GitHub as specified in the corresponding footnotes.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval Not applicable.

[13] https://www.ncbi.nlm.nih.gov/mesh/.

References

1. Banerjee S, Chakravarthi BR, McCrae JP, et al. Large language models for few-shot automatic term extraction. In: Rapp A, Di Caro L, Meziane F, et al., editors. Natural language processing and information systems. Cham: Springer Nature Switzerland; 2024. p. 137–50. https://doi.org/10.1007/978-3-031-70239-6_10.
2. Biggs J, Tang C, Kennedy G. Teaching for quality learning at university. 5th ed. Columbus, UK: McGraw-Hill Education; 2022.
3. Cohen SA. Instructional alignment: searching for a magic bullet. Educ Res. 1987;16(8):16–20.
4. Elmadani M, Mathews M, Mitrovic A. Concept tagging in Moodle. 2012. In: Proceedings of the 1st International Moodle Scientific Conference
5. García-Barragán Á, González Calatayud A, Solarte-Pabón O, et al. GPT for medical entity recognition in Spanish. Multimed Tools Appl. 2024. https://doi.org/10.1007/s11042-024-19209-5.
6. García-González H, Labra Gayo JE, Paule-Ruiz M. Enhancing e-learning content by using semantic web technologies. IEEE Trans Learn Technol. 2017;10(4):544–50. https://doi.org/10.1109/TLT.2016.2629475.
7. Grévisse C, Pavlou MAS, Schneider JG. Docimological quality analysis of LLM-generated multiple choice questions in Computer Science and Medicine. SN Comput Sci. 2024. https://doi.org/10.1007/s42979-024-02963-6.
8. Groza T, Caufield H, Gration D, et al. An evaluation of GPT models for phenotype concept recognition. BMC Med Inform Decis Mak. 2024;24(1):30. https://doi.org/10.1186/s12911-024-02439-w.
9. Grévisse C, Botev J, Rothkugel S. Learning resource management through semantic annotation features in popular authoring software. In: Proceedings of the 10th annual International Conference of Education, Research and Innovation. IATED, ICERI2017, p. 1479–1488. 2017. https://doi.org/10.21125/iceri.2017.0470
10. Manrique R, Grévisse C, Mariño O, et al. Knowledge graph-based core concept identification in learning resources. In: Ichise R, Lecue F, Kawamura T, et al., editors. Semantic technology. Cham: Springer International Publishing; 2018. p. 36–51. https://doi.org/10.1007/978-3-030-04284-4_3.
11. Mazza R, Sommaruga L, Baldassari A, et al (2011) Moodle textclouds. In: MoodleMoot Italia. https://www.moodlemoot.it/file.php/20/Contributi/abstracts/MAZZA-SOMMARUGA-BALDASSARIMAZZOLA__Moodle_TextClouds.pdf
12. Moore S, Schmucker R, Mitchell T, et al. Automated generation and tagging of knowledge components from multiple-choice questions. In: Proceedings of the Eleventh ACM Conference on Learning Scale. Association for Computing Machinery, New York, NY, USA, LS ’24, p. 122–133. 2024. https://doi.org/10.1145/3657604.3662030
13. Reales D, Manrique R, Grévisse C. Core concept identification in educational resources via knowledge graphs and large language models. SN Comput Sci. 2024. https://doi.org/10.1007/s42979-024-03341-y.