

# An educational Gamebook on computational linguistic methods for the development of taxonomies

Fritz Kliche<sup>1</sup> and Ulrich Heid<sup>1</sup> and Ralf Knackstedt<sup>2</sup> and Thomas Klupp<sup>3</sup>

<sup>1</sup>Institute of Information Science and Natural Language Processing

<sup>2</sup>Institute of Business Information Systems

<sup>3</sup>Institute of Creative Writing and Literary Studies

University of Hildesheim

{kliche, heid, ralf.knackstedt, kluppt}@uni-hildesheim.de

## Abstract

We report on a course on computational linguistics and business information systems which includes different concepts of serious games. We developed an interactive Gamebook which features elements such as a contiguous story, quizzes and games. The story mirrors tasks of our students in a laboratory-like part of the course (problem-based learning). In several situations in the story, the readers are given choices for the continuation of the storyline. Based on individual choices, the protagonists in the Gamebook are successful or fail. Wrong decisions anticipate and prevent possible wrong or at least unhelpful decisions in the "real-world" laboratory tasks. We describe elements and concepts of the Gamebook and draw conclusions from an evaluation provided by the course participants.

## 1 Introduction

We report on an ongoing experimental course combining topics from computational linguistics and business information systems which includes serious games, a concept referring to stories, quizzes and games which should be fun and entertaining, but which also have an educational purpose (Bellotti et al., 2013).

The playful elements are offered in an interactive Gamebook which was written specifically for the course. It tells the story of three students attending a fictional university course which roughly covers the same topics as the real-world course. Next to the story, it includes quizzes and games on the contents from computational linguistics and business information systems which we want to convey.

Using the Gamebook, we first aim at a motivational effect provided by playful elements. Second, these elements build on reinforcing teaching strategies rooted in the long traditions of "Programmed Instruction" (e.g. Skinner, 1954 Calleder, 1969) and computer-assisted learning which are based on

the idea that learners profit from immediate feedback.

Another central concept is problem-based learning. Following Boud and Feletti (1997, 2), this strategy does not start with the presentation of knowledge, but with a problem. Knowledge and skills are acquired by a sequence of "problems" which are embedded in a context, supplemented with learning materials and support from the lecturers.

In our course, instead of exercises coined for a specific learning unit, the participants work on a project from the field of business information systems, in order to understand computational linguistic techniques as tools for real-world problems. We are guided by the idea that this approach is similar to applications of NLP methods outside of a classroom situation. They would typically require decisions on appropriate text data and NLP tools, and include the possibility to fail with unsuitable strategies.

The overall goal from the business information systems perspective is the development of domain descriptions by means of taxonomies in the sense of Nickerson et al. (2013). They describe the objects of a given domain, their properties and relations in terms of dimensions, which are attributed with features. During the course – and in the Gamebook – we use a taxonomy on the topic of "carsharing" developed by Schoormann et al. (2017) as an exemplary use case. Table 1 shows three of its dimensions and some attributed features. Overall, this taxonomy has 16 dimensions and 82 features.

Section 2 refers to related work. In section 3, we detail the environment for which the Gamebook was written: The student public and the subject matters we taught. Section 4 reports on the Gamebook and its features. We describe the design for the evaluation of the Gamebook by the course participants in section 5. In section 6, we draw conclusions from the experiences using the Gamebook and the

Dimensions	Features
Vehicle classes	City car   Mid-size car   Van   ...
Customers	Private customer   Business customer   Public sector   ...
Propulsion	Electric   Combustion   Hybrid   ...

Table 1: Part of the taxonomy on carsharing developed by [Schoormann et al. \(2017\)](#).

evaluation and describe our lessons learnt.

## 2 Related Work

Playful elements were often applied for Natural Language Processing. Next to serious games or forms of gamification included in courses, applications also include e.g. games with a purpose (GWAP) which use gamified elements for motivating users to leverage human work intensive tasks such as linguistic annotations, e.g. word sense labeling ([Venhuizen et al., 2013](#)).

The Workshop series "Games and NLP" (cf. e.g. [Madge, 2022](#) for the proceedings of the 9th edition) discusses games and gamification for Natural Language Processing.

[Möslein-Tröppner and Bernhard \(2018\)](#) focus on collaborative aspects of Gamebooks for education. They give best practices for storytelling and for the design of decision paths, e.g. for the structured integration of collaborative elements using flowcharts.

Benefits from gamified features are widely studied. An example of a study on gamified educational tools is provided by [Mazarakis \(2017\)](#). He develops an online quiz with 170 questions on geography and finds a motivational effect when participants acquire "badges" for correct answers, which are icons with e.g. a light bulb symbol, lettering such as "Godlike" or similar motivational content. We adopted this concept for the Gamebook.

[Li et al. \(2020\)](#) describe their experiences during an NLP course using problem-based learning approaches. The task during the course is to develop a text summarization system for a large collection of documents. They evaluated the course with five items on the students' self-assessments. The positive effects on motivation and problem solving ability in the context of the course's topics were higher for undergraduate students than for graduate students.

Motivational and hedonic qualities have also been seen as elements of the evaluation of software, with a focus on user experience and strategies to enhance the motivation of users to work with a given

software, which we could use for our evaluation. E.g., [Hassenzahl et al. \(2000\)](#) investigate the importance of hedonic qualities (e.g., if a software is perceived as interesting), ergonomic qualities (such as ease of use) and the extent to which a software is evaluated as appealing.

## 3 Student public and teaching objectives

### 3.1 Student public

The Gamebook was written for a masters' course with 24 participants at the University of Hildesheim in the summer of 2023.

The course is taught in co-teaching between the business information systems institute and computational linguistics. Consequently, also the student public is diverse: Participants study programs in information systems development, information management, or translation studies and technical writing. Their prior knowledge of computational linguistics and of corpus-based methods is rather limited (maximally one or two BA courses). For this public, getting operational with corpus and NLP tools in a laboratory-like setup is a non-trivial task.

### 3.2 Teaching objectives

The basic idea underlying the course and the Gamebook is that the construction of taxonomies (in the sense of table 1) can be massively supported by computational linguistic tools. Learning objectives are thus (i) taxonomies as an element of information system design, (ii) the design of practical projects to develop a taxonomy by using computational linguistic tools, (iii) the principles underlying the tools and a critical evaluation of their output with a view to taxonomy building.

Following the example of carsharing, the participants of the course are asked to identify a subdomain of the domain of alternative forms of transportation, and to develop a taxonomy for the selected subfield.

The computational linguistic pipeline proposed for taxonomy building involves corpus design (selection of appropriate sources) and corpus develop-

ment, linguistic annotation and corpus exploration. For the latter, not only pattern-based data extraction and querying are offered, but also tools based on BERT architectures (Devlin et al., 2019). Figure 1 depicts the pipeline.

More in detail, concepts of web crawling using the web crawler Trafilatara (Barbaresi, 2021), data cleansing and corpus building are introduced. We address lemmatization, part-of-speech tagging and underlying methods. The corpora are made available in the corpus analysis software CQPweb (Hardie, 2012). We integrate BERTopic (Grootendorst, 2022) for topic modeling and the related keyBERT (Grootendorst, 2020) for keyword extraction. Using a method according to Nickerson et al. (2013) the participants iteratively develop their taxonomies in groups of three students each.

## 4 The Gamebook

The Gamebook is given as an additional source of information for the students, next to standard materials (transparencies, sample data, notes on the principles underlying the tools, as well as on their use).

The Gamebook is divided into six thematic episodes of about 15 pages each, an introduction and an epilogue. It is given in PDF files and is currently implemented as ebook. Students are provided with a new episode every second week, synchronized with the program of the course. The Gamebook distinguished three types of textual elements:

- (A) It contains the story which includes "choose your own adventure" elements: At several points in the storyline readers have to take decisions which impact on the remainder of the episode and in some cases on the overall success of the fictional student team.
- (B) The Gamebook conveys subject-related content on computational linguistics and business information systems. This content is framed by the story, but can be read independently.
- (C) The Gamebook contains several quizzes and games.

### 4.1 Story and decision points

The story is a typical hero's journey, including a quest, beginning in a lecture of a course at the beginning of the semester. The professor talks about

the compilation of taxonomies, and announces that the best group of the course will win a voucher for one year of free carsharing. The story tells how the three protagonists follow the course, sometimes eager to win the competition, sometimes more interested in simply passing the course with little effort, and sometimes failing, depending on decisions of the readers.

While learners read the storyline, they are presented with choices. Depending on their decisions, different continuations of the story are offered (in different sections of the book). Thus, the story is individually adapted for each reader based on the knowledge and mastery of the contents to be learned. At the end of each episode, the different storylines meet again, in order to reduce the number of possible reading paths.

In several parts of the book, readers can collect various forms of points, which are relevant in the final episode: Depending on the results, the end of the story comes in six variants, ranging from a bad and disappointing performance of the three protagonists in their fictional university course, to the best of the possible ends where the fictional characters win the carsharing voucher.

An example for a scene with a decision point is situated at a car exhibition where the fictional students have to go through a sequence of tasks related to corpus exploration strategies, and where each task is waiting next to a given exhibit. The choice of exploration strategy (precision-oriented, providing relatively few examples, most of which are highly taxonomy-relevant vs. recall-oriented, providing much more results, but only a handful of which are relevant for the taxonomy) is connected, for the purpose of the story, with a decision whether a standard family car is to be visited next, or a fancy sports car.

When readers decide for the sports car, the story leads the fictional student group to the rather disappointing result of getting quite few taxonomy-relevant corpus examples, while the other path provides richer and more usable results. In either case, the fictional team verbalizes the reasons for the outcome, stating e.g. that the large amount of un-specific results was due to using the too general query. In some cases, the reader can even decide to go back to an earlier point in the sequence of analysis steps, and to try out an alternative path. Students reading the Gamebook may thus get access to best practice recommendations for their own

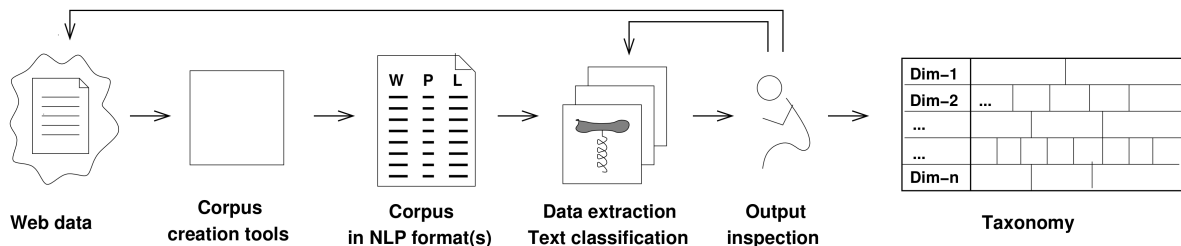


Figure 1: NLP pipeline for the development iterative definement of a taxonomy.

projects without having themselves to lose time in backtracking after avoidable mistakes or unhelpful steps of their work.

## 4.2 Subject-related content

Next to the story, the Gamebook embeds subject-related information. The following quote is an excerpt from the Gamebook and its English translation, giving an example for the embedding of information from the realm of business information systems, here the beginning of a definition of taxonomies.

Du verkneifst dir eine Antwort und packst deine Sachen aus, der Prof schaltet indessen den Beamer ein. „Vor aller Praxis“, sagt er, „kommt aber die Theorie: Wie eigentlich erschließt man sich einen fremden Gegenstandsbereich wie zum Beispiel das Carsharing? Wie rückt man unbekanntem Phänomenen auf den Leib?“ Er sieht erwartungsvoll in den Raum. Dein Sitznachbar – Ben, dieser Statistik-Ben, mit dem Du auch im Mensch-Maschine-Kurs sitzt – meldet sich. „Indem man eine Taxonomie entwickelt“, strebert er los. Der Prof nickt, dann deutet er auf das Whiteboard, auf das der Beamer jetzt einen Text projiziert: „Bitte, zur Einführung!“ Du wischst dir die nassen Haare aus der Stirn und liest:

Taxonomien

- Taxonomien sind Modelle, mit denen Wissen über Phänomene expliziert werden kann. Als Artefakte sind sie vom Menschen geschaffene Werkzeuge.
- Sie dienen den Zwecken, Phänomene anhand von Dimensionen und Dimensionsausprägungen zu beschreiben, zu verstehen, zu analysieren und zu gestalten.

Translated into English:

You refrain from answering and unpack your stuff, while the professor switches on the projector. "Before all practice," he says, "there is theory: How do you make an unfamiliar subject such as carsharing accessible? How do you get to grips with unknown phenomena?" He looks expectantly into the room. The person sitting next to you – Ben, that statistics Ben with whom you also sit in the human-machine course - raises his hand. "By developing a taxonomy," he nerds out. The professor agrees, then he points to the whiteboard onto which the projector is now displaying text: "Please, as an introduction!" You wipe your wet hair out of your forehead, and read:

Taxonomies

- Taxonomies are models that can be used to explicate knowledge about phenomena. As artifacts, they are man-made tools.
- They serve to describe, understand, analyze and design phenomena by means of dimensions and features.

## 4.3 Quizzes and games

We use quizzes and games to allow student readers to test their knowledge of the fields discussed in the course and the decisions they would take in presence of certain kinds of data output from the computational linguistic tools.

As an example, figure 2 shows a part of a game on part-of-speech tagging. The readers have to move through a "board" with  $6 \times 6$  fields.<sup>1</sup> Beginning on the "Start" field, they read the first of 11

<sup>1</sup>In the online version to be developed later in 2023, extra points can be earned by users who get through this parcours particularly quickly.

questions concerning the interpretation of part-of-speech tagging:

Put your game token on the field *Start*. You see output from the part-of-speech tagger. You can reach each field adjacent to *Start*, also fields which are diagonally adjacent. One lemma in the example is wrong. Which one? Move your game token to the corresponding field.

As the sample sentence includes the expression "[erleichtert das] Autoteilen" ("[makes] carsharing [easier]"), the field with the example "Autoteil" is the right choice – as this lemma is wrong.

When readers find the correct field, its adjacent fields contain a correct answer for the second question:

Please search for the underlying tagset online. Are tokenization, part-of-speech tag "KON" and lemma "bzw." correct for token *bzw.*?

The course participants read the Gamebook on their own. We discuss the episodes in the classroom, but it is not checked or controlled which strategies the students chose and which decisions they made while reading. Likewise, the quizzes and games are played individually, and the correct answers are given in the Gamebook itself.

## 5 Evaluation design and early results

We evaluate both the course and the Gamebook with a questionnaire which distinguishes three dimensions:

- One part of the questionnaire deals with the students' expectations and perceptions with respect to the contents and the form of both the course as a whole and the Gamebook. We ask for previous knowledge on the subjects, about the proportions of theoretical parts, practical exercises in the classroom and the time invested for the Gamebook. Concerning the Gamebook itself, we ask for options (5-step Likert scale) on items such as "I enjoy reading the texts" or "there should be more alternative storylines".
- We use 10 questions on students' expectations on self-efficacy, as proposed by [Schwarzer and Jerusalem \(1999, 15\)](#). E.g., the questionnaire asks how the participants evaluate their

problem solving competence in the context of the course.

- We follow the User Experience Questionnaire (UEQ) of [Laugwitz et al. \(2006\)](#) in the reduced form developed by [Alberola et al. \(2018\)](#) consisting of 11 questions. It measures the effectiveness and efficiency, but also the hedonic quality of software products.

A first round of student feedback has been collected in the seventh week of the course as a mid-term evaluation, a second round at the end of the course.

We asked in another text field for a mistake the participants would quite likely had made in their own practice work, had they not first read the Gamebook. Several answers mentioned problems with regard to web crawling. Web sites might not contain enough relevant text data, or the texts collected from the crawler might not be as relevant as expected. Also, the problem of duplicate text content in crawled texts was mentioned.

An interpretation of this feedback could be that the Gamebook is perceived as helpful for the laboratory work of the student groups, but less for theoretical background on computational linguistic methods.

Table 2 shows average results (scale 1-5) from some questions concerning the design of the Gamebook. The evaluation showed that students clearly preferred a realistic scenario (students attending a university class) over e.g. fantasy elements (average of 2.18 resp. 1.71).

The second item asks if the participants read alternative storylines. They indicate with average scores of 2.82 and 2.67 that they rather do not, possibly supporting the assumption that an extrinsic motivation dominates the occupation with the Gamebook.

On the other hand, the participants report that there should be more alternative storylines (average scores of 3.31 and 2.92) – possibly because the decision points were directly related to the tasks of the real-world student projects.

With an average score of 3.53 the students indicate that the Gamebook is rather supportive, and they evaluate it with average scores of 4.31 and 4.27 as creative. Finally, the Gamebook is perceived as rather well understandable (scores 3.82 and 3.71).

The tag is only used for English texts.	It is an error.	No, that is wrong.	Yes
Yes, all tags are correct.	Lemma is wrong.	Bekannte	fallen fällen
partially	Autoteil ←	→ START	Begriff

Figure 2: Example for C. Game on part-of-speech tagging.

Question	Average Score: Mid-term	Average Score: Final
Instead of the setting in a university, I would have liked a story featuring fantasy elements.	2.18	1.71
If there are alternative storylines, I read all of them.	2.82	2.67
There should be more alternative storylines.	3.31	2.92
How do you evaluate the Gamebook: ± supportive	3.53	3.50
How do you evaluate the Gamebook: ± creative	4.31	4.27
How do you evaluate the Gamebook from ± understandable	3.82	3.71

Table 2: Evaluation results referring to the design of the Gamebook.

## 6 Conclusions and lessons learnt

We presented the main components of a Gamebook addressed to a public of master students with little or no background in computational linguistics. The course where we use the Gamebook combines contents from information system design and from NLP. The Gamebook is intended to allow students to get a feeling for best practice use of NLP tools for taxonomy building without having to go through time-consuming and possibly off-putting experiences and mistakes in their own practical laboratory work. We also expect the Gamebook to be an element of motivation.

Based on our evaluation, we draw some first conclusions which might be valuable for similar projects. First, the Gamebook was evaluated as being creative and motivating. Second, we conclude that the story and the playful elements should not deviate too far from the objectives of the lecture. E.g., fantasy elements were not desired according to the evaluation, and elements of theoretical back-

ground which are not directly applicable to the students' projects are evaluated less favorably.

The students confirm that the Gamebook helped prevent pitfalls both with respect to taxonomy building (e.g. mistakes in the taxonomy such as overlapping or redundant dimensions and features) and to the use of corpus data (e.g. duplicate or corrupted content in the crawled text data). This confirms our main motivation for the development of a Gamebook.

The storyline of the Gamebook that follows the model of a quest for a treasure (Möslein-Tröppner and Bernhard, 2018) can be seen as a parable for courses in applied corpus linguistics (research question, corpus design and exploration, presentation of findings). We argue that parts of the fictional story are reusable in different contexts.

Our next steps will be to provide the Gamebook as an ebook made available via an OER portal<sup>2</sup>. Based on the evaluation results, we are interested

<sup>2</sup>www.twillo.de

in detailing the motivational factors of playful elements in both computational linguistics and business information systems, which seem to be most fruitful when they are directly connected to the extrinsic motivation of the course participants.

## 7 Acknowledgements

This work was carried out in the framework of the project "Development of taxonomies", financed by the Ministry of Science and the Arts of the German federal state of Lower Saxony, in the framework of its program "Innovation Plus" (project IP2021/P109).

## References

- Catherine Alberola, Götz Walter, and Henning Brau. 2018. Creation of a short version of the user experience questionnaire UEQ. *i-com*, 17(1):57–64.
- Adrien Barbaresi. 2021. Trafilatura: A web scraping library and command-line tool for text discovery and extraction. In *Proceedings der ACL-IJCNLP 2021*, pages 122–131, Bangkok.
- Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger, and Riccardo Berta. 2013. *Assessment in and of serious games: An overview*. *Advances in Human-Computer Interaction*.
- David Boud and Grahame Feletti. 1997. Changing problem-based learning. introduction to the second edition. In *The challenge of problem-based learning*, pages 1–14, London. Kogan Page.
- Patricia Calleder. 1969. *Programmed learning: Its development and structure*. Longman, London.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. DOI: <https://doi.org/10.5281/zenodo.4461265>.
- Maarten Grootendorst. 2022. Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794.
- Andrew Hardie. 2012. CQPweb: Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380–409.
- Marc Hassenzahl, Axel Platz, Michael Burmester, and Katrin Lehner. 2000. Hedonic and ergonomic quality aspect determine a software's appeal. In *Proceedings of the CHI 2000*, pages 201–208, The Hague.
- Bettina Laugwitz, Martin Schrepp, and Theo Held. 2006. Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In *Mensch und Computer 2006*, pages 125–134, München. De Gruyter.
- Liuqing Li, Jack Geissinger, William A. Ingram, and A. Fox, Edward. 2020. Teaching natural language processing through big data text summarization with problem-based learning. *Data and Information Management*, 1(1):18–43.
- Chris Madge, editor. 2022. *Proceedings of the 9th Workshop on Games and Natural Language Processing within the 13th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France.
- Athanasios Mazarakis. 2017. Gamification: Eine experimentelle Untersuchung der Spielelemente Abzeichen und Story. In *Mensch und Computer 2017 – Tagungsband*, pages 3–14, Regensburg. Gesellschaft für Informatik e. V.
- Bodo Möslein-Tröppner and Willi Bernhard. 2018. *Digitale Gamebooks in der Bildung: Spielerisch lehren und lernen mit interaktiven Stories*. Springer Gabler, Wiesbaden.
- Robert C. Nickerson, Upkar Varshney, and Jan Muntermann. 2013. A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3):336–359.
- Thorsten Schoormann, Dennis Behrens, and Ralf Knackstedt. 2017. Carsharing Geschäftsmodelle: Entwicklung eines bausteinbasierten Modellierungsansatzes. In *Smart Service Engineering: Konzepte und Anwendungsszenarien für die digitale Transformation*, pages 303–325, Wiesbaden. Springer Fachmedien Wiesbaden.
- Ralf Schwarzer and Matthias Jerusalem. 1999. *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Freie Universität Berlin, Berlin.
- Burrhus F. Skinner. 1954. The science of learning and the art of teaching. *Harvard Educational Review*, 24(2):86–97.
- Noortje Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. 2013. Gamification for word sense labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 397–403, Potsdam, Germany.