Lexicon-driven automatic sentence generation for the skills section in a job posting

Vera Aleksić[1], Mona Brems[2], Anna Mathes[1], Theresa Bertele[1]
The Stepstone Group, Munich[1]
The Stepstone Group, Brussels[2]

{vera.aleksic, mona.brems, anna.mathes, theresa.bertele}@stepstone.com

Abstract

This paper presents a sentence generation pipeline as implemented on the online job board Stepstone. The goal is to automatically create a set of sentences for the candidate profile and the task description sections in a job ad, related to a given input skill. They must cover two different "tone of voice" variants in German (Du, Sie), three experience levels (junior, mid, senior), and two optionality values (skill is mandatory or optional/nice to have). The generation process considers the difference between soft skills, natural language competencies and hard skills, as well as more specific sub-categories such as IT skills, programming languages and similar. To create grammatically consistent text, morphosyntactic features from the proprietary skill ontology and lexicon are consulted. The approach is a lexicon-driven generation process that compares all lexical features of the new input skills with the ones already added to the sentence database and creates new sentences according to the corresponding templates.

1 Introduction

Writing and posting a job ad can be time-consuming and expensive, especially for small businesses without a human resources department and with a limited budget. The aim of the Stepstone Recruit project is to accompany the entire recruitment process from the creation and publishing of a new job ad to the matching between the job and the CV database and proposing the best candidates for the vacancy. The process is entirely automated and enables the job publisher / recruiter (hereafter referred to as the user) to create a new job ad within a few minutes, without the need to write any text.

As a smaller but crucial part of this large project, the Linguistic Services team has created a pipeline for the automatic generation of sentences for given input skills to be embedded into the job ad text.

The creation of a job ad starts with the selection of a job title (job descriptor, also referred to as JD). From an auto-suggested list the user can select one of ca. 0.5 million German or English JDs (lemmas) organized in ca. 137.000 concepts (synonym groups) from the proprietary ontology. In the next step, a list of the best matching skills is automatically proposed, to be selected by the user for the given job. The JD-skill matching is powered by an AI model trained and developed in the data science department of the company.

For each chosen skill, a set of pre-generated sentences is provided. The user can either accept the first suggestion or select an alternative, while also being able to edit the wording, as well as to add their own text. The generation of the sentences for ALC (automatic listing creation) is the subject of the current paper.

2 Related work

Researchers are actively exploring methods and techniques to automate and enhance text generation tasks. Recent methods, especially machine learning and AI techniques, have advanced significantly, enabling the development of sophisticated approaches for automatic text generation. Notably, the emergence of models like Generative Pre-trained Transformer, GPT and ChatGPT (OpenAI, 2023) has had a transformative impact on the field of text generation, including

Proceedings of Recent Advances in Natural Language Processing, pages 32–40
Varna, Sep 4–6, 2023
https://doi.org/10.26615/978-954-452-092-2_004

applications in the recruitment domain, such as creating job posting descriptions or cover letters.

In our task, we have adopted a template-based, lexicon-driven approach to sentence creation, carefully designed to accommodate diverse input parameters. To the best of our knowledge, there have been no similar research papers specifically focused on automated generation in the recruitment domain. However, there is a long history of research that combines text mining techniques to extract information from the existing corpus with rule-based NLG methods to generate new text based on specific requirements. Other fields, such as weather forecasting (Saliby, 2019), the financial sector (Pejic Bach et al., 2019), and the medical domain (Hueske-Kraus, 2003) are examples that have received more extensive research and application attention than HR and recruitment. A recent systematic survey (Goyal et al., 2023) provides a comprehensive overview of the history of text generation tools and techniques, shedding light on the evolution of this field and its applications.

3 Corpus analysis and Extraction of text building blocks

To achieve the goal of developing an accurate and domain-relevant system, a corpus of 3.8 million text lines specifying the requirements for the candidate profile (hereafter referred to as the profile section) and 4.8 million text lines containing the job description (hereafter referred to as the task section) was analysed. The corpus was extracted from the job ads in German language published on Stepstone's German board in the past two years.

The profile section usually describes skills, competencies, education, and experience that the candidate needs to have (you have knowledge of <skill>; you have experience in <skill>). The task section describes the tasks and the responsibilities of the role (your tasks will include <skill>, your responsibilities will be <skill>). Both sections were processed separately to analyse the behaviour of different skills in different contexts, as well as to extract relevant patterns per section.

The objective was to generate a collection of phrases and text fragments around a placeholder skill, that could be assembled into complete sentences based on various criteria and parameters outlined below. The second main objective was to generate a set of rules and constraints for the replacement of the skill placeholder with the real input skill. Refer to section 4 for more information. The first version of the system is developed for the German language.

3.1 Skill ontology

To ensure syntactically and morphologically accurate replacement of the skill placeholder, all relevant information about the skill must be coded in a lexicon. This includes information such as the base form of the skill, its inflected forms, and any additional semantic or classification codes.

A rich domain ontology representing occupations, skills, industries, education qualifications and other key concepts in the recruitment domain has been the core of all relevant processes in the semantic search, classification, normalization, matching, recommendation, and analytics in the organisation. In the beginning of the project, the skill sub-ontology contained about 80.000 concepts (semantic clusters) with almost 215.000 lemmas (synonyms and translations in a few languages), of which around 62.000 lemmas in German. In addition to the semantic and ontological information, each lemma is stored with its inflected forms and corresponding morphosyntactic features. Minimum required information is part of speech, gender, and number. The corresponding linguistic rules are maintained in an inflection module. All this information serves as input parameters to the sentence generation pipeline. If insufficient, the ontology allows the feature sets to be extended with additional semantic or pragmatic codes at the concept or lemma level. The coding/tagging functionality was extensively used to iteratively optimise the lexicon for the purposes of the project. In the following sections, more details will be given.

While one part of the improvements and annotations remains project-related and language-specific (German), another key by-product of the tagging process is the development of a more generic language-independent skill classification schema, applied to all skills at the concept level.

3.2 Skills in context and concordances

In the previously mentioned corpus, all German skills from the ontology were identified within the text, and concordances were generated to show their surrounding context. For this, the tool Unitex (Paumier, 2008) was used. As the user manual of the tool says: "Syntactic graphs, often called local

grammars, allow you to describe syntactic patterns that can then be searched in the texts. Of all kinds of graphs these have the greatest expressive power because they allow you to refer to information in dictionaries." (Paumier, 2008).
In compliance with the requirements of the tool, the skills were exported into the DELA (Dictionnaires Electroniques du LADL) format (Courtois, 1990) containing inflected forms and semantic codes (Figure 1).

```
belagsmaschinenführung,belagsmaschinenführung.N+ACT+HARDSKILL:ns:gs:ds:a
belagsmaschinenführungen,belagsmaschinenführung.N+ACT+HARDSKILL:np:gp:dp
belagsverlegung,belagsverlegung.N+ACT+HARDSKILL:ns:gs:ds:as
belagsverlegungen,belagsverlegung.N+ACT+HARDSKILL:np:gp:dp:ap
belarussischem,belarussisch.A+ACT+LANGUAGE:MdsIRTP:NdsIRTP
belarussischen,belarussisch.A+ACT+LANGUAGE:FapEATP:FdpEATP:FdpIRTP:FdsEA
belarussische,belarussisch.A+ACT+LANGUAGE:FapIRTP:FasEATP:FasIATP:FasIRT
belarussisch,belarussisch.A+ACT+LANGUAGE:Uf:XP
belarussisches,belarussisch.A+ACT+LANGUAGE:NasIATP:NasIRTP:NnsIATP:NnsIR
belastbarkeitstest,belastbarkeitstest.N+ACT+HARDSKILL:Mas:Mds:Mns
belastbarkeitstests,belastbarkeitstest.N+ACT+HARDSKILL:Map:Mdp:Mgp:Mgs:M
belastungsfähigem,belastungsfähig.A+ACT+PERSONAL:MdsIRTP:NdsIRTP
belastungsfähigen,belastungsfähig.A+ACT+PERSONAL:FapEATP:FdpEATP:FdpIRTP
belastungsfähige,belastungsfähig.A+ACT+PERSONAL:FapIRTP:FasEATP:FasIATP:
belastungsfähig,belastungsfähig.A+ACT+PERSONAL:Uf:XP
```

Figure 1: DELA dictionary, compiled for this project

The following semantic codes were established for the initial analysis of the corpus:

- **soft skill** with the subcategories: *personal* (self-confident), *social* (team-oriented), *methodological* (attention to detail)
- **hard skill** with these two special sub-categories: *IT skill* (computer skills and tools such as Java, Cloud Security, UX design), *tool* (tools used in production, skilled trades, logistics etc.: Abrasive wheels, Truck cranes, Blueprint machines)
- **language skill** (knowledge of natural languages: German, English)

The codes were added to the skills at the concept level. Skills that were not tagged in the ontology are considered hard skills without any additional classification.

In Unitex, a simple local grammar (Figure 2) was created containing a single skill box without specifying any surrounding context. To further refine the analysis process, the grammar was divided into individual sub-graphs, with each sub-graph dedicated to a specific skill class as defined in the ontology and in the DELA dictionary.

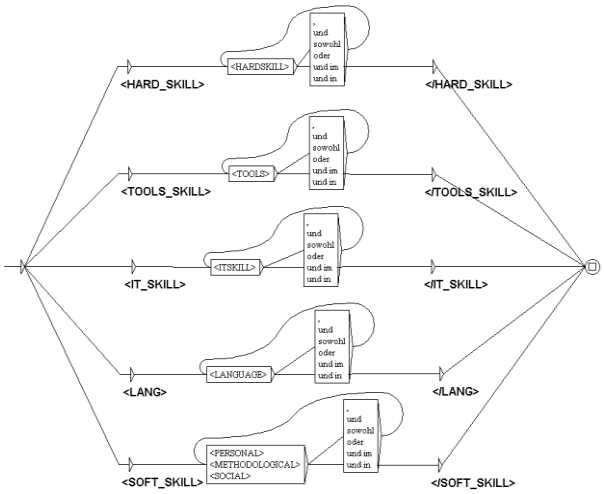

Figure 2: Syntax Graph

The skill boxes in the syntax graph as presented in Figure 2 match any form of any lemma as defined in the lexicon. When multiple skills are concatenated with a comma or a coordinating conjunction, they are grouped within a single <skill> tag to better distinguish the boundaries between the central skill position and its preceding and following context within the concordance.

In Unitex, the chosen length for the left and right context in the concordance view was set to 60 characters each, so the lines in the result file were not equal to sentences. Using a sentence splitting approach for German as described by Thurmair (2012), the potential beginning of a sentence in the left context and the potential ending of a sentence in the right context were detected, and anything outside of this scope was deleted to produce cleaner text for further processing.
Lines that were clearly wrong in the profile or in the task section were deleted. The same went for the lines without any useful content and, of course, for lines without recognised skills in them. The result of this task was a list of approx. 1 million lines.

3.3 Properties of the profile section and the task section

First tests and analyses confirmed the difference between the text structure in the profile section and task section.

In the profile section, the original general distinction between IT skills, tools and other hard skills was not fully mirrored in the found patterns. For some of the analysis tasks, it was sufficient to operate on a generic <skill> placeholder, instead of the tagged skill blocks as in the concordances

(hard_skill, tools_skill, it_skill) since most of them followed the same patterns and were embedded into similar contexts. On the other hand, it was discovered that for certain groups of skills a distinction on the lemma level, rather than the class or concept level, would be necessary. For example, instead of only 'IT skill', programming languages need to be classified separately (experience in developing in <Python>); also IT tools follow other patterns (experience in working with <Adobe Framemaker>); in cases when the skill lemma already contains the activity (developing, programming), it needs to be embedded into different patterns (<Programmieren in Java>, <Umgang mit MS Office>) ("<programming in Java>, <working with MS Office>").

Languages and soft skills have clearly shown that they behave differently in the context. Here as well, it was discovered that additional classification is needed according to the lemma form and its morphology rather than according to the class or concept. The following three lemmas belong to the same language skill, but follow different patterns depending on the part of speech and the noun form (for illustration, we will use comparable English examples):

- You speak the <German language> very well.
- You have very good <German skills>.
- You have a very good knowledge of <German>.

Some examples of the classification/annotation codes at the lemma level will be listed further below in section 5.6.

One notable finding from the analysis is the scarcity of soft skills within the task section. They are observed to be rarely mentioned or represented in this section. While the portion of soft skills in the profile section was 26,1%, in the task section it was only 12,8%. Some of the skills were short, semantically incomplete, and very generic (coordination, presentation) and sometimes false positives as soft skills. During the project many such skills were extended or replaced in the ontology by more appropriate skills, extracted from the relevant context found in the concordances (good presentation skills; coordination of manufacturing processes).

Furthermore, it was observed that in the task section, descriptions involving language skills often include multiple skills. For example, phrases such as "your job will include translation from German into English" imply the need for proficiency in both German and English. However, the current project was initially designed to process only one skill as input. As a result, the handling of multiple input skills was deferred to future development stages.

3.4 Application input parameters

The final selection of patterns and preparation of text building blocks was also determined by the following business requirements:

- The input parameter that is referred to as "**tone-of-voice**" should allow the user to choose how to address the job seeker, either by using the polite address in German (Sie), the informal address (Du), or to rather use an impersonal form (n/a).

- Another input parameter is the required **level of experience** for a particular skill, which the user can select to be junior, mid, or senior. If the level of experience is not applicable (e.g., for the soft skills) the input value for the text generation will be n/a. For languages, the value "native" also exists.

- The third parameter is the **optionality** of the given skill, which can be selected as either mandatory (true) or optional/nice to have (false)

    - To ensure diversity and an optimal exposure of various alternatives to the user, the main pragmatic business requirement for each skill is to generate a **minimum of three distinct sentences for each legal combination of tone-of-voice, experience, and optionality**.

The analysis revealed that certain skills may not be compatible with all combinations of the three input parameters. Consequently, the following restrictions were applied to the different types of skills in the profile and in the task sections:

- In the profile section, hard skills and language skills have no restrictions on tone-of-voice, level of experience, or optionality. This means that text with all possible combinations of these three parameters can be generated.

- Soft skills occur frequently in the profile section but have no level of experience.

- Soft skills are typically not used in the task section at all.
- Language skills in the task section will be omitted for now, as described above.
- Consequently, in the task section, only text containing hard skills needs to be generated. Hard skills in the task section are always mandatory.

4 Selection of syntactic components, compilation of lexical resources, and development of government rules

The frequent pre-context and post-context blocks around the <skill> position were divided into smaller chunks (Figure 5) according to their part of speech and their role in the sentence (Rothstein, 2008).

Subject	Verb	Prep	Adj	Noun	Prep		Particle
Sie	haben		erste	Kenntnisse	in	<skill>	
Sie	bringen		fundierte	Erfahrungen	im	<skill>	mit
Du	verfügst	über	solides	Wissen	im	<skill>	

Figure 5: Fine-grained sentence chunks

This was the basis not only for the extraction of the syntactic patterns, but also for the collection of the most frequent lexical resources to fill the subject/predicate/object positions, including the corresponding adjectives, adverbs, prepositions, and conjunctions.

4.1 Seed sentences and iterative improvements

By identifying and selecting a first set of reliable patterns and corresponding lexical fillers, it was possible to generate a corpus of seed sentences. Examples were compiled for each semantic and syntactic category, with skill positions being filled by lexicon entries that matched the slot constraints. Through manual evaluation of approximately 11.000 sentences across three iterations, a set of generation rules was developed and iteratively refined. In addition, new semantic and syntactic codes were introduced and applied to existing lexicon entries, and standardised processes were established for tagging all new and future entries in the ontology.

4.2 Syntactic components and word order

A set of fundamental syntactic structures was selected to be used as templates for the text generation Here are some basic examples:

SVO (subject-verb-object) word order:

1. Du sprichst gut <Deutsch>
   "You speak <German> well"
2. Du bist ein <Team Player>
   "You are a <team player>"
3. <Englisch> ist Deine Muttersprache
   "<English> is your mother tongue"
4. Du bringst [Wissen über <Java>] mit
   "You bring [knowledge of <Java>] with you]"
5. Du hast [Erfahrungen in <Java>]
   "You have [experience in <Java>]"
6. [Erfahrungen in <Java>] wären ein Plus
   "[Experience in <Java>]) would be a plus"
7. [Dass Du <Deutsch> sprichst], ist ein Plus
   "[Speaking <German>] is a plus"
8. Was Dich auszeichnet ist, [dass Du <Deutsch> sprichst]
   „What sets you apart is [that you speak <German>]"

VSO (verb-subject-object) word order:

9. Vorzugsweise sprichst Du <Deutsch>
   "You preferably speak <German>"
10. Idealerweise bist Du ein <Team Player>
    „Ideally you are a <team player>"

Depending on the verb and its syntactic valency, and on the word order, the role of the <skill> in the clause can be: **direct object** (after transitive verb, like in examples 1, 9); **subject complement** (after copula verb like in examples 2, 10); **subject** (example 3); **prepositional phrase complement** in a noun phrase with heads such as "experience" or "knowledge", which function either as a direct object (examples 4, 5) or a subject (example 6); **object in a subordinate clause**, which can function as the subject of the main clause (example 7) or as the subject complement of the main clause (example 8).
To accurately populate the skill slot, its nominative form is selected for the subject, the accusative case for the direct object, and in prepositional phrases the choice between the dative or accusative is governed by the preposition.

4.3 Syntactic government rules

To guide the assembly of the single components into text, a set of syntactic government rules was derived and manually enhanced. Also, further rules were established to ensure the correct morphological agreement between the skill and its associated syntactic structures, including verb conjugation, article usage, declension of adjectives, and word order. Here are some rules:

- Main word order is SVO. The order will change to VSO if the optionality adverb takes the first position in the sentence. The finite verb in German remains in its default second position, since all generated sentences are declarative.

- The choice of a tone-of-voice value must be considered when the personal pronoun (**Du, Sie**) ("you") is the subject of the clause, or its accusative form is the object of the clause (**Dich**, **Sie**) ("you"), also if a possessive pronoun is a part of the pattern (**Dein** Profil, **Ihr** Profil) ("your profile"), and for the correct generation of the subject-verb agreement (Du **verfügst**, Sie **verfügen**) ("you have").

  The particle of a separable verb is placed at the end of the main clause (Sie bringen Erfahrungen in Machine Learning **mit**) ("You bring experience in machine learning").

- In subordinate clauses, the particle is not separated (Was Sie **mit**bringen, sind Erfahrungen in Machine Learning) ("What you bring is machine learning experience").

- In general, subordinating conjunctions move the final verb to the end of the clause (Was Sie vorweisen **können**, sind Erfahrungen in Machine Learning) ("What you can show is experience in machine learning").

- The number of the input skill governs the number of the verb (Java **gehört** (sg) zu Ihren Stärken. Java-Kenntnisse **gehören** (pl) zu Ihren Stärken) („Java belongs (sg) to your strengths. Java skills belong (pl) to your strengths").

- Articles are declined depending on the gender and number of the input skill, and of the case required by the preposition (Erfahrung **im** Management. Erfahrung **in der** Programmierung) ("Experience in management / in programming").

- If the skill is the subject complement with a copula verb, it needs an indefinite article (Sie sind **ein** Team Player) ("You are a team player").

- Some soft skills (but not all) require an indefinite article if they are subject to the verb "have". Those skills are tagged in the ontology (**ein** Organisationstalent) ("a talent in organisation").

4.4 Casing

Another set of rules is used to adapt the casing in the sentence. In the ontology, all lemmas are capitalised, independently of their part of speech. In the text, adjectives and verbs must be lowercased. For better readability, skills in an apposition are enclosed in double quotes, and remain capitalised (You have experience in the field of "**T**echnical acoustics"). Nominalised verbs remain capitalised in the sentence.

4.5 Lexical resources

For each of the sentence components a list of alternative expressions was created, as for example:

- The predicate position is filled with different verbs (have, posses, bring along, master etc.), or with the copula verb "be". Some of them need a tag in the lexicon.

- Experience levels can be expressed in many ways, such as with adjectives (junior: basic, mid: solid, senior: extensive), with the length of experience (many years), or by using idiomatic expressions (You are a true master in Java).

- If the skill is mandatory for the position, this can be expressed either by the present indicative form of the verb in the main sentence (you have, you bring along), or by idiomatic phrases (Java knowledge is a must). Optional skills can be expressed

by adverbs (ideally, optionally), or by idiomatic phrases (is a plus).

- To avoid string repetition, some patterns are excluded from the choice if the skill itself contains the same (sub)string. (**Kenntnisse** in <Java-**Kenntnissen>,** im **Bereich** <Finanz**bereich>**) („knowledge in <knowledge of Java>, in the domain of <finance domain>“).

The lists of alternative words and phrases include the most frequently occurring expressions extracted from the corpus, ensuring comprehensive coverage of commonly used language variants.

4.6 Ontology annotations

In some cases, the information created in the lexicon by the standard inflection and annotation modules is not sufficient to meet all requirements for the sentence generation process. Several new codes were introduced at the lemma level. In the table below are some examples of tag assignments pertaining to syntactical as well as to semantic properties:

Tag	Example skill	(English)
indef_art	Auge fürs Detail	Eye for detail
base_lang	Afrikaans	Afrikaans
lang_knowledge	Igbo-Kenntnisse	Knowledge of Igbo
adj_lang	Arabische Sprache	Arabic language
verb_lang	Deutsch sprechen	To speak German
adj_substlang	Südliches Sotho	Southern Sotho
subst_substlang	Khmer-Sprache	Khmer langauge
be	Team Player	Team player
prog_lang	C++	C++
work_with	SAP	SAP
show	Eigeninitiative	Initiative
have	Ausdauer	Endurance
neg_connotation	Bankbetrug	Bank fraud

Table 1: Examples of lemma tags

Tags are used to for example dictate the usage of indefinite articles (indef_art) or the usage of certain verbs as predicates (be, have, show). Semantic tags such as neg_connotation prevent generation of phrases that require experience in illnesses, fraud, terror acts or similar. Instead of “you have experience in money laundering” other formulations are taken, to rather indicate experience as a specialist in this domain. Language skills have many different tags used to select the correct predicate in the clause (speak, know, have (knowledge)).

5 Automatic process in production

Subsequently, the final collection of rules and resources was automatically applied to all the skills featured in the ontology, resulting in the creation of almost 30 million unique sentences for approximately 62.000 skills.

To streamline the regular production process, a pipeline was established to generate sentences for newly added skills in the ontology. The infrastructure incorporates the databases for the ontology maintenance and storage of the sentence data, a Python pipeline for the generation of new sentences, and a serverless process for automatic export of new data to the production system.

5.1 Generation pipeline

The sentence generation pipeline comprises the following sequential steps:

- The pipeline begins by examining the ontology for any newly added skills. A temporary dictionary is created.
- The features of the new skills, such as concept tags, lemma tags, number, and gender, are compared against previously processed skills.
- In the case of multiword phrases, the head positions are compared, and the pipeline searches for the longest common ending of the head tokens. If a common length of at least 3 characters is found, the skill is immediately selected as an example for the new skill.
- If not all requirements are met, the rules are relaxed. This includes accepting shorter common endings and allowing a fewer number of matching tags. These relaxations are logged for further manual checks and analysis.
- The pipeline selects the best matching example skill based on the previous comparisons.
- Corresponding example sentences are retrieved from the database, serving as templates for generating new sentences.

5.2 Quality assurance

The formal automatic quality assurance task encompasses the evaluation of the following aspects: number of patterns per skill, number of sentences per pattern, deletion of skills from the ontology (it checks if any skills were removed from the ontology, to delete the corresponding data sets in the sentence DB).
In addition to the formal automatic evaluation, a qualitative evaluation is conducted through manual checks of the logs. This evaluation aims to identify the reasons for missing example skills or insufficient coverage. Potential reasons could include missing tags, a wrongly assigned head position for the lemma, incorrectly inflected forms, or wrong features. Typically, improvements in the ontology are required to address these issues and enhance the quality of sentence generation.

6 First results

In the pilot phase of the project, 20 stakeholders were asked to test the system, to use it to publish their job ads, and to give their explicit feedback.

In total, 465 listings (job ads) were published. Quantitatively, the published listings contained a total of 2031 sentences in the profile sections. 1289 of them were taken from the auto-suggest without any modification. The task section contained 1580 sentences in total. 1229 of them were taken from the auto-suggest without any changes.

All other sentences were added by the job publisher during the job ad creation. In both sections, it was observed that most of the lines that did not come from the auto-suggestion option were rather simple enumeration, either single skills (mostly free-text, so out-of-vocabulary skills) or skill lists, where similar skills were grouped together in one bullet point.

Qualitatively, the stakeholders' descriptive feedback was positive. They appreciated that the job ad could be created in very short time. In the first test round, the sentences were sometimes perceived as schematic and uniform, with similar structure and same sentence beginnings. To overcome that, an external module was developed to select sentences with the longest lexical distance.

The following qualitative feedback provided by stakeholders and internal testers significantly influenced the further development of the generation module:

SVO sentences with personal pronouns "Du" or "Sie" as the subject, were perceived as dominant and monotonous. Instead, neutral sentences where the skill itself served as the subject (Several years of experience in the field of "Virtual Design" are a must) or sentences with as subordinate clause as the subject (What you have already acquired is basic experience in output management) were considered more natural and appealing. As a result, the number of such sentences was increased.

Sentences expressing that the skill is optional were found to be richer and more varied compared to sentences with mandatory skills. This is because many of the "nice to have" patterns follow the VSO order, with the optionality adverb typically taking the first position in the sentence, and as such they offer an opportunity to enhance the diversity of the sentence beginnings by expanding the vocabulary for the given slot (e.g. ideally, desirable, optional, advantageous).

Since the selection of the formal, informal, or neutral tone as well as the specification of optionality are essential input parameters, the freedom in selecting slot fillers is limited. To introduce more variety, strategies such as sentence order inversion and the use of idiomatic fillers (e.g. <skill> is a must; a must is <skill>) were employed.

7 Summary and Outlook

The paper describes a method for automatic creation of content using pre-established rules and templates, without any reliance an artificial intelligence or machine learning algorithms. The process enables to quickly create high-quality and contextually appropriate content. It offers numerous advantages including efficiency, consistency, customisability, and accuracy. It can serve as both the primary approach for sentence creation and as fallback option for AI methods.

The qualitative improvements will concentrate on allowing multiple skills as the input to the generation.

Lastly, user feedback will be integrated into the development loops (free-text skills that are not in the ontology yet, modified sentences, patterns that never were selected or were discarded by the user).

References

Mirjana Pejic Bach, Živko Krstić, Sanja Seljan, and Lejla Turulja. 2019. Text Mining for Big Data Analysis in Financial Sector: A Literature Review. *Sustainability*, 11. 1277. 10.3390/su11051277.

Federica Calanca, Luiza Sayfullina, Lara Minkus, Claudia Wagner, and Eric Malmi. 2019. Responsible team players wanted: An analysis of soft skill requirements in job advertisements. *EPJ Data Science*, 8(1). doi:10.1140/epjds/s13688-019-0190-z.

Blandine Courtois. 1990. *Un système de dictionnaires électroniques pour les mots simples du français*. Larousse, Langue Française 87, Paris.

Ehtesham Ul Haq Dar and Jürgen Dorn. 2018. *Ontology based classification system for online job offers*, 1-8. 10.1109/ICOMET.2018.8346340.

Yunus Doğan, Feriştah Dalkılıç, Alp Kut et al. 2020. Discovering the same job ads expressed with the different sentences by using hybrid clustering algorithms. *International Journal of Applied Mathematics Electronics and Computers*, 8(3), pp. 76–84. doi:10.18100/ijamec.797572.

Katja Filippova and Michael Strube. 2007. Generating constituent order in German clauses. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. Prague, Czech Republic, pp. 320–327.

Rupali Goyal, Parteek Kumar and V.P. Singh. 2023. A Systematic survey on automated text generation tools and techniques: application, evaluation, and challenges. *Multimedia Tools and Applications*, 1-56. 10.1007/s11042-023-15224-0.

Akshay Gugnani and Hemant Misra. 2020. Implicit skills extraction using document embedding and its use in job recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(08), 13286–13293. doi:10.1609/aaai.v34i08.7038

Dirk Hueske-Kraus. 2003. Text Generation in Clinical Medicine – a Review. *Methods of Information in Medicine*, 42, pp. 51-60. 10.1055/s-0038-1634209.

Malgorzata Mochol, Radoslaw Oldakowski, and Ralf Heese. 2004. *Ontology based Recruitment Process*. https://www.researchgate.net/publication/2913821_Ontology_based_Recruitment_Process#fullTextFileContent

Renate Musan. 2021. *Satzgliedanalyse*. Universitätsverlag Winter. Heidelberg, Germany

OpenAI. 2023. ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

Karin Pittner and Judith Berman. 2021. *Deutsche Syntax Ein Arbeitsbuch.* Narr Francke Attempto. Tübingen, Germany

Sébastien Paumier. 2008. Unitex 2.0. User Manual. Université Paris-Est Marne-la-Vallée, Paris

Björn Rothstein. 2018. Syntax – die Analyse des Satzes und seiner Bestandteile. In: Dipper, S., Klabunde, R., Mihatsch, W. (eds) *Linguistik.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-55589-7_3

Joe G. Saliby. 2019. Survey on Natural Language Generation. *International Journal of Trend in Scientific Research and Development*. Volume-3, 618-622. 10.31142/ijtsrd22903.

Gregor Thurmair, Vera Aleksić and Christoph Schwarz. 2012. Large-scale lexical analysis. *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, LREC 2012, Istanbul, Turkey

Michael Tschuggnall, Benjamin Murauer, Günther Specht, and Julia Brandl. 2018. *Algorithmic Segmentation of Job Ads Using Textual Analysis*. https://www.researchgate.net/publication/359857179_Algorithmic_Segmentation_of_Job_Ads_Using_Textual_Analysis#fullTextFileContent

Hans Uszkoreit. 2023. *Word Order and Constituent Structure in German*. Bibliovault OAI Repository, the University of Chicago Press.