DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

When voice assistants take hold: How smart technology makes goods receipt more efficient

Wenn Sprachassistenz Einzug hält: Wie smarte Technologie die Effizienz im Wareneingang erhöht

Heiner Ludwig
Mathias Kühn
Thorsten Schmidt

Professur für Technische Logistik
Institut für Technische Logistik und Arbeitssysteme
Fakultät Maschinenwesen
Technische Universität Dresden

Goods receipt is an essential step in supply chains. It is often performed manually by warehouse staff. They are confronted with challenges such as time pressure, missing goods, wrong quantities, damages and incomplete documents. To support the workers in the best possible way, digital assistance systems are suitable. The study aims to investigate the potential of voice assistance in various processes of goods receipt. These include verbal identification of incoming goods and interactive guidance in quality control. Technical requirements will be defined, an adaptive data structure will be presented, and a novel neural network will be trained for multiple inputs. Finally, two example scenarios are evaluated and compared with currently used methods and tools.

[Goods receipt, Voice assistant, Natural language processing]

Der Wareneingang stellt einen essentiellen Schritt in Lieferketten dar. Dieser wird dabei oftmals von Lageristen manuell durchgeführt, die mit unterschiedlichen Herausforderungen wie Zeitdruck, fehlenden Waren, falsche Mengen, Schäden und unvollständigen Dokumenten konfrontiert sind. Um die Arbeitskräfte bestmöglich zu unterstützen, sind digitale Assistenzsysteme geeignet. In der Studie soll das Potenzial der Sprachassistenz in verschiedenen Prozessen des Wareneingangs untersucht werden. Dazu gehören die verbale Identifikation eingehender Güter und die interaktive Anleitung in der Qualitätskontrolle. Es werden technische Anforderungen definiert, eine lernfähige Datenstruktur vorgestellt und ein neuartiges neuronales Netz für Mehrfacheingaben trainiert. Abschließend werden zwei Beispielszenarien evaluiert und mit derzeitig genutzten Methoden und Werkzeugen verglichen.

[Wareneingang, Sprachassistent, Natürliche Sprachdatenverarbeitung]

1 Introduction

Goods receipt is an important, recurring process in supply chains. The data recorded about incoming goods plays an essential role in warehouse management and ensures transparency along the supply chain. Unreported, untreated defects may continue in the value chain and create error chains. Incoming goods are mainly handled by warehouse clerks, who will continue to play an important role in the near future due to their versatility and flexibility. However, especially in high-wage countries, the workers are a major cost factor. In order to increase the speed and accuracy of incoming goods processes, to document and rectify deviations quickly and to avoid media disruptions, it makes sense to optimally support warehouse staff with digital assistance systems.

We are investigating the use of mobile, voice-based assistance systems, which offer clear advantages, such as the high interaction speed and the hand- and eyes-free operability for parallel use during the actual manual work process [1]. An efficient voice-based interaction approach in the form of pick-by-voice systems was already established decades ago and ensures faster work with low distraction in the picking process [2]. While pick-by-voice systems are based on a small command vocabulary and a number system, voice assistants with advanced natural language processing (NLP) have become commonplace, guaranteeing high accessibility and processing of complex data inputs (cf. *Amazon Alexa*, *Google Assistant*) [3].

The aim of this study is to test the suitability of a mobile voice assistance system for assisting warehouse clerks in goods receipt. For this purpose, we show typical processes and difficulties as well as currently used aids in goods receipt and derive application scenarios and requirements for the voice assistant from them. This concerns the verbal data recording by the user for feedback in the process of incoming goods (e.g. delivery lists, container types, quantities, detailed documentation of defects) as well as an interactive guidance by the system, e.g. when taking samples by inexperienced warehouse staff. In addition, we propose a suitable data structure and train a neural network that subdivides the documentation of several incoming goods from one speech input. Finally, we carry out an evaluation for each scenario to compare the effectiveness of the voice assistant in goods receipt with currently used tools and methods.

2 State of the art in research and practice

For an overview of the status quo, a literature review was carried out and interviews were conducted with various companies. In doing so, we noticed that processes in the incoming goods area have been little studied so far, yet numerous problems occur in practice. [4] traces strong variances in the lead time of medical supply chains back to goods receipt and sorting processes and points out that errors in goods receipt can have fatal consequences in the medical sector. [5] extract process chains to be completed in goods receipt and compare different problems in terms of the probability of their occurrence and their effects.

To elaborate on the general approach to goods receipt and to uncover problems and challenges, systematizing expert interviews by [6] were conducted in each case. We interviewed five companies from different industries and gathered their processes, challenges and equipment used. Among them is a distribution center of a large supermarket chain, a medium-sized company for logistics consulting, a manufacturer and supplier of photovoltaic systems, an aircraft manufacturer and a railway construction company with loading personnel between 8 and 60 persons per shift. Information was gathered on the respective general organization and process in goods receipt, challenges for warehouse staff, de-escalation strategies in case of deviations and the usage of digital tools.

2.1 Process steps

The receipt of goods is usually carried out according to a fixed procedure that is specified by legislation and standardization and is similar across industries (cf. [7, 8, 9]). Depending on the receiving goods and the situation, the time available for the processes, the number of random samples and the measures in error handling vary. Furthermore, additional legal requirements may have to be taken into account, e.g. for food products (cf. [10, 11]). For a better overview, we have shown a schematic sequence of a goods receipt in Figure 1. This is based on [5] and the processes of the interviewed companies.

2.2 Tools used

Various tools are used for the documentation of incoming goods. In some cases, printed packing and check lists are filled out by hand and then transferred to a digital warehouse or supply chain management system. For the most part, handheld devices with a touch-based input are used. They either have pen-based operation via resistive touch surfaces or work finger-based with a capacitive surface. Some of them are equipped with an additional scanner and/or have an additional physical keyboard. Various industry partners use glove scanners that can be used hands-free. Furthermore, terminal PCs are sometimes used, which are installed on trolleys and driven to the loading ramp as required.

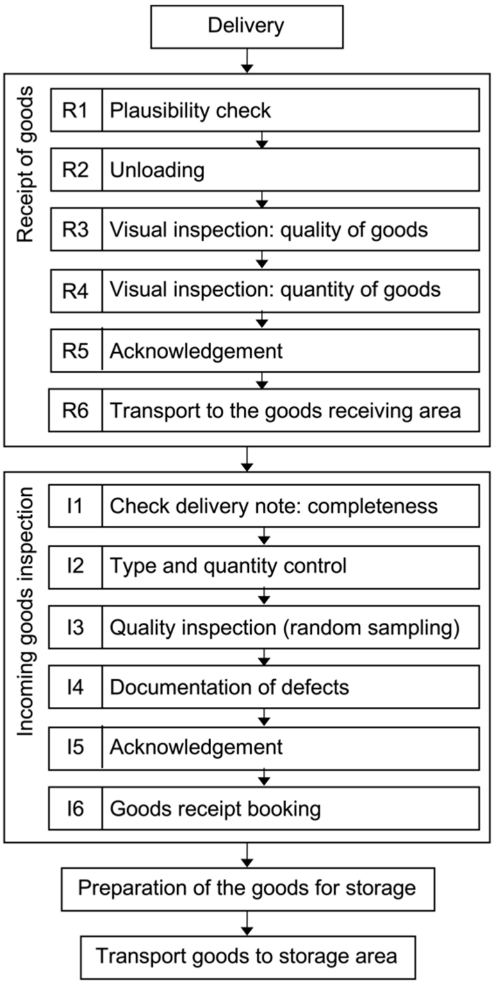
Delivery
Receipt of goods
R1 Plausibility check
R2 Unloading
R3 Visual inspection: quality of goods
R4 Visual inspection: quantity of goods
R5 Acknowledgement
R6 Transport to the goods receiving area
Incoming goods inspection
I1 Check delivery note: completeness
I2 Type and quantity control
I3 Quality inspection (random sampling)
I4 Documentation of defects
I5 Acknowledgement
I6 Goods receipt booking
Preparation of the goods for storage
Transport goods to storage area

Figure 1. Schematic sequence of a goods receipt

New types of systems and research projects are attempting to automatically record incoming goods using RFID chips or image processing methods. This is done either via mobile or fixed positioned devices. RFID gates are already commercially available and continue to be a focus of attention in the scientific community [12, 13]. The main advantage is the high recognition speed, especially at stacked goods. On the other hand, there is the additional cost of consistently equipping all containers with RFID chips along the supply chain. In addition, the visual inspection for damage to the goods still has to be carried out. In this field, there are several studies on how image data processing can be used for automatic quality control for specific cases [14, 15]. However, both approaches only cover partial areas of goods receipt and require a high initial effort and cost to deploy. They are to be regarded as supplementary systems that take over sub-processes from the warehouse operator, but do not replace him completely.

An initial, coarse granular quality check in goods receipt is usually done by the loading staff before a more detailed check is carried out by corresponding experts if necessary (e.g. in the case of specially manufactured components that have certain functional requirements or must comply with prescribed standards). For this, the warehouse operator must be trained in the various test procedures and steps. Often this is done through instructions by experienced staff, the use of checklists and the provision of printed manuals. There are also various software applications for documentation that are used on mobile or stationary devices. Particularly in the case of large quantities of identical goods, image processing systems are also occasionally used to evaluate the samples. However, these have so far only been suitable for certain scenarios and small expected product assortments and require high initial investment costs (cf. [16, 17], *Siemens Vision System SIMATIC MV*).

2.3 Challenges for the warehouse operator

The challenges in goods receipt for the warehouse operator are numerous. We have summarized 9 challenges from [5] and own company interviews (cf. section 2). The delivery of consumer products as well as industrial goods were considered. Further requirements and fields of application for the voice assistant are to be derived from this. While the first challenge is the smooth handling of the receiving process by the warehouse clerk, the subsequent challenges are related to the handling of deviations from the standard process. For each challenge, the affected processes from Figure 1 are indicated in the right header cell.

<table>
<tr><th>#</th><th><i>Challenge</i></th><th><i>Affected processes</i></th></tr>
<tr><td><i>C1</i></td><td><i>Fast, complete receipt of goods</i></td><td><i>R1, R3, R4, I1, I2, I3, I4, I6</i></td></tr>
<tr><td colspan="3">The basic task is to accept the goods quickly and carefully, otherwise delays, error chains and additional work will occur in following processes. This applies to both delivery processes (e.g. accumulation, additional demands) and warehouse processes (e.g. subsequent notification of damaged goods, asynchronous digital data storage).</td></tr>
</table>

<table>
<tr><td><i>C2</i></td><td><i>Receiving an incorrect quantity of goods</i></td><td><i>R1, R4, R5, I1, I2</i></td></tr>
<tr><td colspan="3">A deviating number of incoming goods includes both missing and over-delivered goods. It has to be clarified whether the difference to the target quantity occurred during or already before the transportation. If no deviation is known after consultation with the supplier, complete documentation of the previous delivery processes is required to find the deviation (see C3). If the quantities are lower than expected, the reasons must be documented. These are partly pre-categorized and must be selected by the employee. Surplus goods are either not accepted in principle or accepted after consultation with the purchasing department. Due to different working hours of the incoming goods department and the purchasing department, communication problems occur (e.g. during night shifts).</td></tr>
</table>

<table>
<tr><td><i>C3</i></td><td><i>Receipt of goods with incomplete or poorly completed delivery notes and labels</i></td><td><i>R4, R5, I1, I2, I5</i></td></tr>
<tr><td colspan="3">Due to carelessness, lack of experience, damaging, frequent repacking and time pressure, incomplete or unreadable delivery notes and container labels occur. A lack of real-time data transfer between heterogeneous warehouse management systems can also lead to recorded data not being fully available.</td></tr>
</table>

<table>
<tr><td><i>C4</i></td><td><i>Receiving damaged goods</i></td><td><i>R3, I3, I4, I5</i></td></tr>
<tr><td colspan="3">Damage can occur either to the container or to the contents. Damaged contents may not be immediately apparent and are only discovered during spot checks. In the case of infrequently occurring faults and special goods, it is sometimes unclear what steps need to be taken. Thus, it has to be assessed whether the goods are completely rejected and returned, or partially accepted and damaged items are destroyed in the warehouse. If (especially inexperienced) employees are uncertain, the shift supervisor is often contacted. This person is usually also responsible for the manual inventory adjustment in the digital warehouse system.</td></tr>
</table>

<table>
<tr><td><i>C5</i></td><td><i>Damage to goods during their receipt into the warehouse</i></td><td><i>R6, I4, I5</i></td></tr>
<tr><td colspan="3">Goods can be damaged during transport from the loading bay to the storage location in the warehouse due to improper transport or poor packaging. Depending on the extent of the damage, the goods are repackaged or disposed of. This is usually agreed verbally with a responsible clerk or inventory manager. This person documents the damage, adjusts the inventory data accordingly and ensures the disposal and cleaning of the area. In addition, a repeat order is processed via the purchasing department.</td></tr>
</table>

C6	*Incorrect delivery*	*R1*
In rare cases, the complete batch delivered is incorrectly addressed and is not accepted. This impedes the schedule in goods receipt, especially with tightly timed windows. The responsibility for finding the correct recipient lies with the supplier, so that no further action by the warehouse clerk is necessary.		

C7	*Incorrect identification of goods*	*R1, R4, I1, I2*
Incorrect identification can have several reasons. For example, goods may have been transported in the wrong container, incorrect barcodes may have been assigned, or incorrect barcodes may have been scanned (e.g. those of individual products as opposed to those of the container). Visual inspection is necessary to detect the faults, as otherwise error chains that are difficult to trace arise unnoticed.		

C8	*Wrong sorting, wrong packing size*	*I1, I2*
Different container sizes and sorting of goods cause delays in goods receipt, as new comparisons with packing lists have to be made and repacking is necessary. In some cases, incorrect container sizes are not accepted and are returned immediately; this is decided on a situational basis by the purchasing department.		

C9	*Sampling inspection varies depending on the goods*	*I3, I4*
Mixed assortments of goods make it necessary to take different procedures and inspection criteria into account during the sampling inspection. Differently qualified groups of warehouse staff are formed for this purpose. For example, in the receiving area of supermarket goods, a distinction is made between dry goods, fresh and frozen goods and fruit. Fresh goods and fruit or custom components in particular are subject to special inspection procedures for quality assurance and determination of origin. In some cases, personnel from a separate quality inspection department are also added, e.g. for the functional testing of special technical components.		

Based on the challenges, we derive two deployment scenarios for a voice assistant to support staff in goods receipt: challenges C1, C3 and C7 describe the problems of identifying incoming goods correctly and the consequences of incomplete, delayed receipt of goods. The focus here is on the fast and correct data recording of incoming goods. For this, we investigate the challenges and suitability of voice-based goods entries for digital data recording. Regardless of the use of digital systems and tools used today, numerous personnel agreements are required between different departments when unintentional deviations occur. Challenges C2, C4, C5, C8 and C9 lead to more complex decisions by the warehouse staff and to queries from supervisors. This creates a high degree of staff dependency, which has a negative impact on the process in case of non-attendance (e.g. due to different shift times, absence due to illness). In addition, sampling, especially for goods that rarely repeat (such as customized components or promotional goods), often requires experienced, specially qualified workers who are not always available (e.g. due to high turnover rates [18]). For this purpose, we are testing the use of a voice assistant as a digital system for action guidance or support of the warehouse clerk.

3 Method

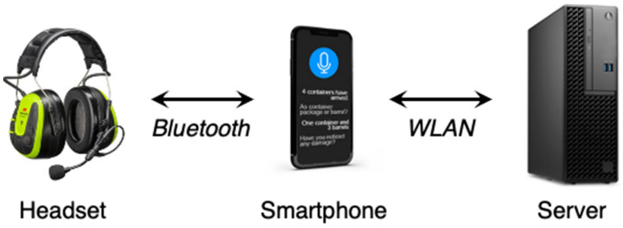

Figure 2. Communication between the system components

We build on the process steps and challenges mentioned in section 2 to provide a voice assistance system for the warehouse worker throughout the goods receipt process. The system consists of three components (cf. Figure 2): a noise-cancelling headset, a smartphone and a server. All devices are wirelessly connected to each other. The headset minimizes recorded ambient noise from delivery and loading vehicles, as well as from loading personnel working in parallel, so that the system correctly recognizes the spoken words. The server is used both for converting the voice input into text form and for central data storage in the company. All data is processed internally in the company network, an internet connection is not required for the operation of the system. The warehouse clerk can then start the dialogue hands-free with a keyword/wake word ("Hey Warehouse, […]"). The speech input is converted into text by using *Open AI's* free speech-to-text engine *Whisper* [19] and evaluated in terms of content using the modelled dialogues and data structures on the server side. In case of unclear data entries or for giving further instructions, synthetic voice data output is generated.

3.1 Data structure

The data structure organizes the database for the overall system in form of an ontology. This includes delivery data from different suppliers and recipients, such as batch compositions, delivery times, products and defects already reported. The system imports and maps data from the warehouse management system, relies on a manually created data structure, or captures data from external digital sources (cf. Figure 3). The digital data sources depend on the product category. For consumer goods, for example,

DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

product databases using EAN numbers are suitable, individual order and construction data provide additional information about unique industrial goods like individual products and customized components. We use this enrichment with meta-data to identify the goods in the goods receipt process by means of natural speech input and ensures greater transparency. The data structure successively expands by the automatic interpretation of the voice inputs in the goods receipt. This concerns documentation of incoming goods and containers, especially with regard to type, quantity, time of arrival and newly occurred damage. Terminology management plays an important role in assessing whether a piece of information already exists in a different formulation or needs to be added. Therefore, after each extension of the ontology, the system automatically checks and adds synonyms from predefined thesauri if necessary. Since the system evaluates all information directly on site, additional travel, media disruptions and personnel dependencies are minimized.

Various studies already exist in other domains on the extension of ontologies with the help of language models, but using texts, not spoken information [20, 21, 22]. We use context-free word embeddings from the pre-trained german *FastText* language model to extract previously unknown information from speech input and insert it into the ontology [23]. Word embeddings are numerical representations of single words in the form of vectors that capture their semantic meaning based on their co-occurrence in text corpora. We determine the similarity measure by the cosine distance of the respective word embeddings from the speech input to each word embedding within the ontology to correctly insert new information from the input language data and ignore irrelevant information. The larger the distance between the word embeddings, the more similar the compared words are. If the value falls below 0, it can be assumed that the words having an opposite meaning are not relevant.

First, we remove all irrelevant terms (e.g. articles, personal pronouns, etc.; so-called "stop words") from the speech input and sort out the known terms by matching them with the existing data (and their synonyms) from the ontology. In the second step we compare unknown words, i.e. new information, for similarity with existing ontology entities to evaluate relevance and classify them thematically. For example, if the warehouse clerk reports specific defects for the first time, the system automatically groups them as subclasses of the "damage" class based on the similarity of the word embeddings, while it discards irrelevant words if there is no similar entity in the ontology.

$$\cos(\theta) = \frac{A \cdot B}{\| A \| \| B \|}$$

Equation 1. Cosine distance between vector A and vector B

3.2 Voice-based identification system for incoming goods

Normally, numbering systems clearly describe the containers, and these are transferred to barcodes if necessary. In logistics, pick-by-voice systems use shorter, simplified numbering systems for simple, fast voice input. Particularly long numbering systems are unintuitive for human speech, prone to errors and can thus lead to a high expenditure of time. For this reason, we present three adapted identification systems in the following:

S1 Numbering system: We evaluate the use of a numbering system as it is already used in practice. We have created exemplary 13-digit EAN styled identification numbers. Since it is predictable that time-consuming, technical character strings require long speech inputs, this shall serve as a point of comparison of the following identification systems.

S2 Symbol based system: We added a data set of 100 simple symbols for a pictorial, centesimal coding identification system. This is to test whether inputs can be made faster, less mentally stressful and with less risk of confusion by using natural symbols. In addition, the system can be used barrier-free. When selecting the symbols, we paid attention to unambiguity and clarity. Not all symbols are suitable for loud pronunciation; the use of a "fire" symbol, for example, is not appropriate because of possible misinterpretations among neighboring employees when speaking in. Since

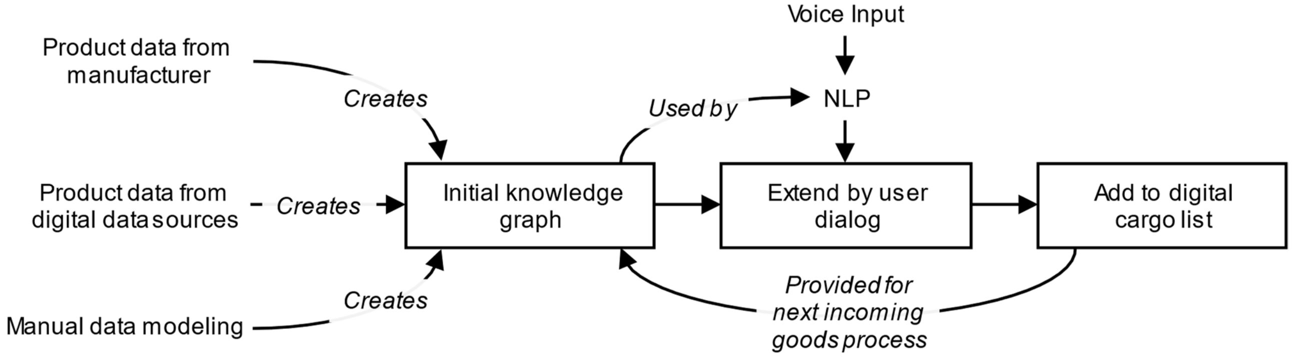
Voice Input
Product data from manufacturer
Creates
Used by
NLP
Product data from digital data sources
Creates
Initial knowledge graph
Extend by user dialog
Add to digital cargo list
Creates
Manual data modeling
Provided for next incoming goods process

Figure 3. Schematic representation of knowledge graph generation

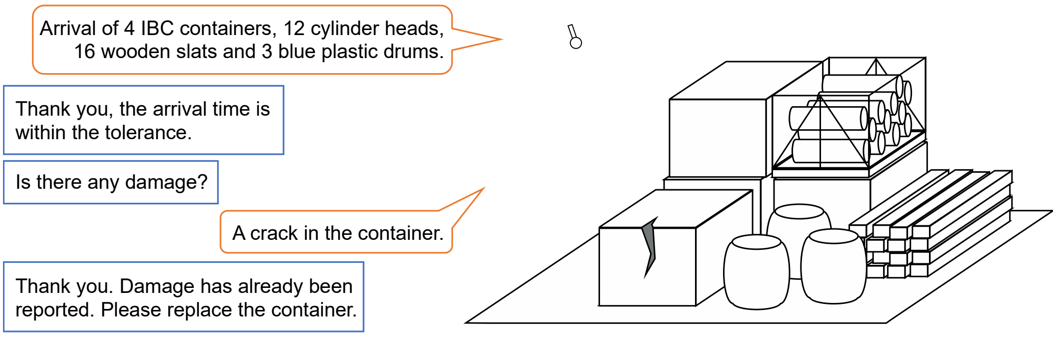
Arrival of 4 IBC containers, 12 cylinder heads, 16 wooden slats and 3 blue plastic drums.
Thank you, the arrival time is within the tolerance.
Is there any damage?
A crack in the container.
Thank you. Damage has already been reported. Please replace the container.

Figure 4. Exemplary dialogue in natural language in goods receipt

different associations may occur with the symbols (e.g. "low shoe", "men's shoe" or simply "shoe", 4th symbol from the left in Image 1), we created synonymous designations in the underlying data structure.

S3 Word based system: Thirdly, we use a system to convert the string of numbers into individual words. We take inspiration from *What3words* [24] and translate the respective groups of numbers into sentences. While with What3words the three words describing a geographical coordinate have no context, we form real sentences to increase the recognition rate by the speech-to-text engine of the voice assistant. The engine is based on probabilistic analysis methods and works more precisely with word associations in frequently occurring word groups. In addition, the word combinations should be as foreign to the product domain as possible so that the voice assistant does not misinterpret the input.

S4 Use of metadata: This system is fundamentally different from the previous systems and is not a classic coding system. An attribute system is defined in the data structure that stores metadata of the goods with regard to external appearance, delivery address, destination and content. Supported by an exclusion mechanism, goods can thus already be identified during the visual inspection on the basis of natural language input without having to read or scan labels. In this way, missing or illegible codes can be counteracted (cf. challenge C3, Section 2.3). This system is intuitive to use and adapted to natural language, but is poorly suited to similar-looking packages with missing labels.

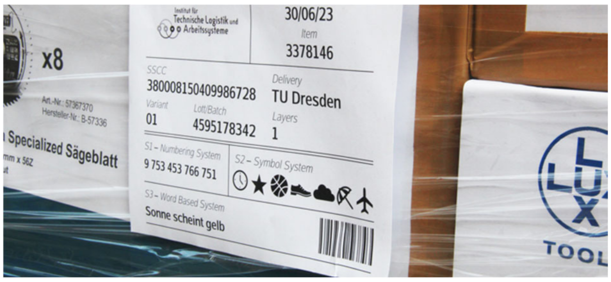

Image 1. Package label with identification systems S1, S2, S3

3.3 Batch data processing

A special feature in goods receipt is the recording of multiple data with one voice input in the incoming goods inspection. While the inclusion of unambiguous data (cf. S1, S2, S3 of Section 3.2) that are independent of each other can be detected by simple matching procedures, the grouping of natural descriptions is more complex (cf. S4 of Section 3.2; Figure 4). A three-step process groups the speech inputs, then checks each for unambiguity, and finally verifies the completeness of the information. This "chunking"- procedure is comparable to the multi-intent-detection mechanisms, which are an important field of research in NLP [25]. However, there are differences in the handling of incoming goods. While multi-intent-detections aim at distinguishing different intents in a speech input, in the goods receipt there are similar intents to record multiple goods and quantities. We first tested the grammatical dependency parser from the open-source library *spaCy* [26], which is based on a non-monotonic arc-eager transition system [27]. The dependency tree created analyses the language input so that related information is grouped together. However, there are weaknesses, e.g. the inaccurate assignment of demonstrative pronouns and generally in the correct evaluation of colloquial sentence structures. For this reason, we have trained a sequential neuronal network which groups the voice input text based on the contextual word embeddings.

3.3.1 Generation of the training data set

Since no suitable training data set exists, we created a separate data set for goods receipt. To accelerate the manual data generation, we developed an application that displays randomly composed goods receipts with different, schematically displayed container types that have to be described by voice input (Figure 4). In the process, fixed attributes (e.g. color, text, content) are to be named, which are fully stored in a database. After data input, we subdivide the recorded text into attribute groups on the basis of the stored attributes. We then used text augmentation to automatically increase the amount of training data based on the manually recorded data. The original dataset contains 457 labelled entries and was enriched to 50 000 entries using text augmentation. Both datasets are published on

GitLab[1]. We used methods of *Rule-based augmentation* and *MixUp augmentation* [28]. This includes *random swapping* and *random deletion* of words to simulate incomplete and grammatically unclean speech input. *Random synonym replacement* is normally used to increase the variety of sentences by replacing individual words. Since the goal of the neural network is to group sentence parts and not to process the content of the sentences, the words do not necessarily have to be synonymous; it is sufficient if the word type matches.

3.3.2 Network architecture

We converted the words of the grouped sentences into word embeddings using a pre-trained German BERT language model2 and combined them into a $W_{max}\ x\ n_{WordEmbedding}$-dimensional matrix as input data for a sequential neural network, where $W_{\max}$ describes the highest number of words in a sentence in the training dataset and $n_{WordEmbedding}$ the number of dimensions of the word embedding vector (BERT encodes word vectors in 768 dimensions by default). The output data is a W_{max}-dimensional vector. The coding is based on the sentence position of the words, consecutive words with the same numbering belong together. We trained two networks with different outputs: one with alternating numbering between 1 and 2 and one with ascending numbering. If a sentence has less words than W_{max}, the remaining positions are filled with 0. The values in the output vector v are floating point numbers and are rounded to a vector v' of integers. The more reliably the neural network evaluates a word from the sentence input, the closer the corresponding output value is to an integer. Figure 5 shows the complete chunking process of both networks. The chunking step corresponds to the respective alternating or ascending trained network. Accordingly, the shape of the respective rounded output vector is $v'_{\mathrm{alternating}}$ or $v'_{\mathrm{ascending}}$.

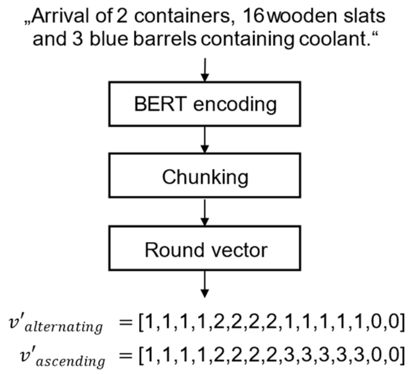
„Arrival of 2 containers, 16 wooden slats and 3 blue barrels containing coolant.“
BERT encoding
Chunking
Round vector
$v'_{alternating}$ = [1,1,1,1,2,2,2,2,1,1,1,1,1,0,0]
$v'_{ascending}$ = [1,1,1,1,2,2,2,2,3,3,3,3,3,0,0]

Figure 5. Chunking procedure for $W_{max} := 15$

A sequential neural network with 8 fully connected hidden layers, each with 64 neurons and the reLu activation function, and a W_{max}-dimensional output layer performs best for the ascending input data. For alternating training data, the same configuration with 3 fully connected hidden layers produces the best results. Between layers, we insert dropout layers with a dropout rate of 0.15 to avoid overfitting. The mean square error (MSE) serves as the loss function and the stochastic gradient descent (SGD) as the optimization function of the neural networks.

3.3.3 Performance evaluation

To evaluate the performance of the neural network we use the following three metrics M1, M2 and M3. We list the results in Table 1.

M1 Number of correctly identified groups: The number of correctly identified groups in the output vector $|G_{v'_{i_{true}}}|$ is set in relation to the number of all groups $|G_{d_i}|$ from datapoint d_i. $v'_{i_{true}}$ describes the partial vector of v'_i which contains the values for correctly identified groups, $v'_{i_{false}}$ is the partial vector with incorrect predictions. A group G is considered to be correctly predicted if all words that belong together match exactly the specified group in the test data. The arithmetic mean is formed over all data points d.

$$M1 = \frac{1}{d}\sum_{i=1}^{d} \frac{|G_{v'_{i_{true}}}|}{|G_{d_i}|}$$

Equation 2. Evaluation metric M1

M2 Number of correctly identified groups without stop words: In addition to important information, such as attributes for a more detailed description of the incoming goods, the recorded data also contains words that are irrelevant in terms of content and cannot be clearly assigned to any group. Conjunctions such as "and" for linking parts of sentences cannot be clearly assigned and are not relevant for goods identification. *M2* is constructed similarly to *M1*. We remove the stop words from the test data set and the chunking prediction, thus forming the groups *G'*, so that the evaluation of the prediction is not biased by irrelevant words.

$$M2 = \frac{1}{d}\sum_{i=1}^{d} \frac{|G'_{v'_{i_{true}}}|}{|G'_{d_i}|}$$

Equation 3. Evaluation metric M2

[1] *https://tlscm.mw.tu-dresden.de/scm/repo/git/VO4.0_IncomingGoods*

[2] *https://huggingface.co/bert-base-german-cased*

M3 Certainty of the prediction of the network: To assess the certainty of the neural network, we use the unrounded prediction vectors. As already mentioned, the certainty of the prediction increases as the difference between the floating point values and the rounded integer values of the output vector decreases. We calculate the difference vector between v and v' for the correctly ($M3_{true}$) and incorrectly/not recognized groups ($M3_{false}$) separately. We then form the magnitude of the difference vector and divide it by the number of vector dimensions to calculate the average deviation. This is repeated for all d and the arithmetic mean is calculated.

$$M3_{true} = \frac{1}{d}\sum_{i=1}^{d} \frac{\sqrt{(v_{i_{true}} - v'_{i_{true}})^2}}{W_{max}}$$

Equation 4. Evaluation metric M3: correctly identified groups

$$M3_{false} = \frac{1}{d}\sum_{i=1}^{d} \frac{\sqrt{(v_{i_{false}} - v'_{i_{false}})^2}}{W_{max}}$$

Equation 5. Evaluation metric M3: incorrectly/not identified groups

We compare four neural network configurations using the three metrics presented. We trained two networks with alternating and two with ascending output vectors, one network with the origin dataset d_{origin} containing 457 data points and one network extended with the augmented training dataset d_{augm} containing 50 000 data points each.

Table 1. Chunking performance for $|G| \leq 3$

Metric	Chunking network			
	Alternating		Ascending	
	d_{origin}	$d_{origin} \cup d_{augm}$	d_{origin}	$d_{origin} \cup d_{augm}$
M1	0.06	0.69	0.12	0.75
M2	0.18	0.77	0.32	0.8
M3$_{true}$	0.19	0.08	0.23	0.13
M3$_{false}$	0.16	0.48	0.79	0.61

Table 1 shows the significantly better performance of the networks trained on larger datasets. The certainty of the networks is significantly higher for the correctly identified groups $M3_{true}$than for the incorrectly identified groups $M3_{false}$. Thus, further processing of misidentified groups can be avoided using M3. As mentioned at the beginning, we use the network as a complementary mechanism, since the performance is not yet sufficient to use it alone. In particular, we see potential to increase performance in an extended rounding procedure, e.g., using smoothing procedures that filter the noise of the output vector.

3.4 Reasoning system

The automatic inference, based on spoken information and a pre-defined set of rules, enables data to be inferred without being explicitly reported by the warehouse clerk. For this purpose, we defined simple logical rules in the data structure, which allow inferring or exclusion of information. Figure 6 shows an example of a restriction for chemical goods that liquid basic materials are only transported in IBC containers. If the warehouse clerk reports "The isopropyl has arrived" as a voice input, the reasoning system infers the container type (an *IBC container*) based on the *hasPermittedContainer* data property, therefore no system-side query to the speaker is necessary. This shortens the dialogue duration and makes the interaction appear more natural. We use the default "open-world-assumption", which means that non-existent information does not automatically mean that a statement can be considered false, but unknown [29]. In the context of the ontology shown in Figure 6, this means that bulk containers are suitable for transporting chemical solids, but this does not preclude IBC containers from being unsuitable (unless a negated *hasPermittedContainer* relation between *Solid* and *IBC container* is added).

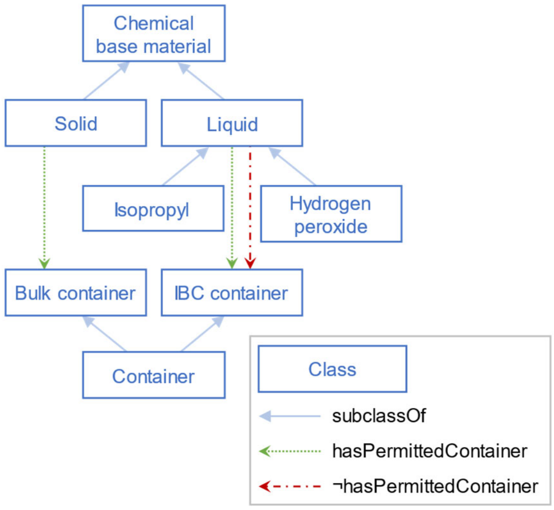

Figure 6. Example ontology for the delivery of chemicals with container restriction

3.5 Virtual assistance

In addition to pure data recording, we are also investigating the suitability of the voice assistant for instructing the warehouse clerk. Our dialog system as virtual assistant minimizes staff dependency and can be used directly on site, avoiding long communication chains. In addition, the data modelling of the dialogues persists knowledge and experience values, which can also be used for training purposes. We focus on type, quantity and quality control in the

DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

incoming goods department (cf. processes I2, I3, I4 in Figure 1). Depending on the product, the warehouse operator must check a variety of properties that often require extensive expertise (cf. Section 2.3, challenge C9). For this purpose, we have implemented a generalized, step-by-step dialogue system with which quality criteria can be defined on an assortment- and product-specific basis. The dialogue components are also stored in an ontological data structure and can be created without code knowledge and easily reused in this way. For example, standards to be checked can be modelled once as dialog in the ontology and then modified/reused to form new dialogues for different products.

The information to be entered is first typed. Thus, the dialogue structure determines whether numeric, binary, predefined or free response options are possible. If the user response deviates from the predefined types, the system asks for a new voice input and clarifies expected types. Triggers can be defined for dynamic conversation, which activate certain dialogues based on user input. If, for example, the tolerance limit is exceeded for a queried measured value, further instructions for action can be output to the warehouse clerk. In addition, a "help" instruction is stored for each step, which provides more detailed information on the execution of the respective test step when requested by the speaker.

4 Evaluation

We investigate the suitability of the voice assistant in goods receipt using two example scenarios (cf. Section 2.3). The goal in each case is to measure the time required for comparison with currently used tools and systems and to identify advantages and challenges in different situations. We conducted the evaluation at a loading dock of a warehouse on the campus of the TU Dresden. For this purpose, we set up pallets with packaged goods, with which a goods receipt is exemplified. The evaluation was performed with an ambient volume level between 48db and 63db.

4.1 Comparison of identification systems

We compared the identification systems described in Section 3.2 in terms of time required. As a reference value, we performed a run with a mobile scanner. It should be noted here that in practice, especially with longer supply chains, uniform barcodes are not always available or are not readable due to package damage. This is taken into account in the evaluation in that the numbers for one container have to be entered manually on the scanner's touch display. We examine the delivery of various auxiliary materials (cf.

Image 2). Here, several goods are packed together and provided with a label that contains the different identification systems S1-S3 (cf. Image 1). Four test persons carried out the evaluation. For this, we have four runs each for the different identification systems for three differently packed pallets with one, four and eight goods respectively.

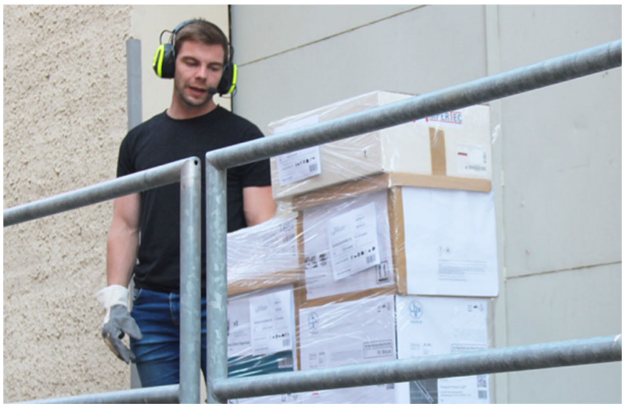

Image 2. Test setup for identification systems

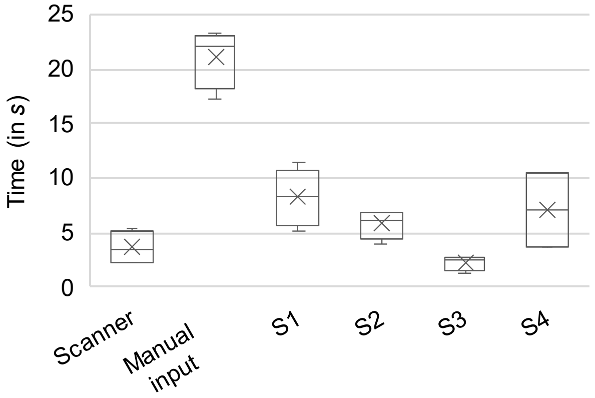

Figure 7. Time required to identify one incoming good

Amount of goods		Scanner	S1	S2	S3	S4
	1	3,6s	8,2s	5,7s	2,2s	4,78s
	4	18,5s	37,2s	29,9s	10,2s	14,8s
	8	37,9s	76,7s	61,5s	23,1s	28,2s

Figure 8. Avg. time required to identify a complete delivery with multiple goods

The test showed that S3 is the fastest verbal data acquisition with the lowest deviations. S1 and S2 turned out to be cognitively stressful. One test person noted that "positive" symbols (e.g., "sun") were more pleasant to speak than "negative" symbols (e.g., "rain") in S2. For S4, the higher deviation is due to the learning effect during the evaluation. Several test persons initially described the incoming goods very extensively, but noticed during the tests that simple descriptions were sufficient for identification. We have also observed that with S4, there are benefits in terms of work ergonomics, as the warehouse operator does not have to bend or stretch to scan or read low/high positioned labels.

DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

4.2 Digital assistant for quality control

As noted in Challenge C9 in Section 2.3, experienced personnel, who may be specially qualified, are required to perform certain sampling procedures. We are investigating to what extent inexperienced staff can be guided step-by-step through a multi-part inspection process by the support of a virtual assistant. As a test scenario, we have chosen the sampling-based quality control of hollow bars made of plastic, as can classically occur in a mechanical engineering company. The interactive virtual assistant guides the worker step by step through the various test steps and gives instructions for action according to the user's responses. The quality control comprises 9 test steps and 7 interactive action instructions depending on the error case. First, the type and number of items is checked and the packaging and products are inspected for integrity. The test person takes random samples and tests various dimensions (outer and inner diameter, rod length). For this purpose, the system uses tolerance tables according to ISO 2768-1 in order to evaluate deviations and issue instructions for action by the worker [30]. When prompted by the "help" command, the virtual assistant gives the worker more detailed information on how to carry out the current test step.

We performed three test runs with four test persons two times each: *Run 1* with an intact sample, *Run 2* with a sample whose material differs from the order and *Run 3* with a damaged sample sent to the quarantine warehouse for further examination. We carry out the inspection process once with voice-assisted and once with an analog inspection protocol, on which passed quality criteria and damage must be entered by hand. For the analog test protocol, a short manual is provided for a more detailed description of the quality control steps. Each test person initially practiced running through a quality control without temporal recording.

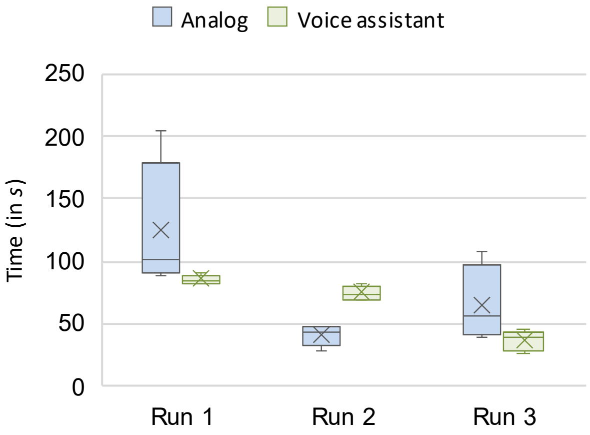

Figure 9. Time required for quality control of one sample

The temporal evaluation shows that the voice assistant of Run 1 has a significantly lower temporal deviation than the analog test protocol. This is due to the continuous dialog guidance, which leads through the quality control using a guide and completely records all information. In Run 2, this is a disadvantage, since information is recorded first that is not relevant in the case of the wrong material. Here, the extension of the dialog system is necessary, which allows a more dynamic dialog flow. In Run 3, the warehouse operator benefits from the input speed of verbal error data recording compared to handwritten notes, especially for more extensive documentation. Not taken into account is the time required to transfer the handwritten notes (e.g. to provide the QM department with documentation of the damage). This step takes place automatically using the voice assistant.

5 Conclusion

In this paper, we highlight the current problems and challenges in manually performed goods receipt. From these problems we derive the two scenarios *Receipt of goods* and *Incoming goods inspection* and investigate the usability of a mobile voice assistant as an easy to learn, fast to use tool. For this purpose, we propose an adaptive data structure in the form of an ontology that is built manually, uses different data sources, or voice input. In the first scenario, we compare four verbal identification systems of incoming goods with a typical scanning process. We also train a neural network that is used to subdivide content sentence components and supports multiple inputs in a single speech data input. In addition, the reasoning system infers further information based on the stored data structure. In the second scenario, we implement a step-by-step sampling quality control system that runs through several check steps in form of a dialog with the warehouse clerk and offers instructions for action in the event of an error. Here we compare the voice assistant with typically used analog test protocols and manuals in terms of time required.

The evaluation of the scenarios demonstrates that voice assistance is an adequate tool to support various processes in the incoming goods department. The overall system works particularly well for goods receipts that regularly receive large quantities of different types of products. The evaluation shows the advantages in terms of time and work ergonomics compared to the hand-held scanner. Hands-free voice interaction to record information directly during the unloading process also provides a time advantage. The evaluation of the quality control process shows that the step-by-step implementation enables even inexperienced warehouse workers to perform a complete random inspection. For faster use by more experienced employees, however, the dialog system has to be more dynamic so that several inspection stages can be run through by one speech input. Further evaluations are necessary for the continuous expansion of the system; in particular, performance needs to be tested with multilingual loading personnel.

DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

Literature

[1] S. Wellsandt, K. Klein, K. Hribernik, M. Lewandowski, A. Bousdekis, G. Mentzas und K.-D. Thoben, „Towards Using Digital Intelligent Assistants to Put Humans in the Loop of Predictive Maintenance Systems,“ *IFAC-PapersOnLine,* pp. 49-54, 2021.

[2] N. Dujmesic, I. Bajor und T. Rozic, „Warehouse Processes Improvement by Pick by Voice Technology,“ *Tehnički vjesnik – Technical Gazette (TV-TG), Volume 25, Issue 4,* pp. 1227-1233, 2018.

[3] Business Wire, Voicebot.ai, „Anzahl verwendeter Sprachassistenten weltweit 2019 und eine Prognose für die Jahre 2020 und 2024 (in Milliarden),“ 2020. [Online]. Available: https://de.statista.com/statistik/daten/studie/1323879/umfrage/anzahl-weltweit-verwendeter-sprachassistenten/.

[4] C. Kiilu, „The Influence of Goods Receipt and Sorting and Loading and Delivery to Client Practices on Inventory Management Performance in Humanitarian Organizations; A Case Study of Afghan Red Crescent Society Medical Supply Chain,“ *International Journal of Supply Chain Management,* p. 62–74, 2016.

[5] E. Kulinska und J. Giera, „Identification and Analysis of Risk Factors in the Process of Receiving Goods into the Warehouse,“ *Foundations of Management, Volume 11,* pp. 103-118, 2019.

[6] C. Helfferich, „Leitfaden- und Experteninterviews,“ in *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden, Springer Fachmedien Wiesbaden, 2019, pp. 669-686.

[7] VDI Verein Deutscher Ingenieure e.V., „VDI 3612 - Wareneingang/Warenausgang,“ *VDI-Handbuch Materialfluß und Fördertechnik, Band 7,* pp. 1-29, 1996.

[8] DIN e.V., „DIN EN ISO 9001:2015-11. Qualitätsmanagementsysteme-Anforderungen,“ 2015.

[9] Bundesamt für Justiz, „Handelsgesetzbuch § 377,“ *Bürgerliches Gesetzbuch (BGB),* 2015.

[10] Food Safety Management Systems, „FSSC 22002-1,“ *FSSC 22000, v6.0,* 2023.

[11] IFS Management GmbH, „Standard for auditing product and process compliance in relation to food safety and quality,“ *IFS Food Standard,* 2023.

[12] H. Knapp und G. Romagnoli, „RFID systems optimisation through the use of a new RFID network planning algorithm to support the design of receiving gates,“ *J Intell Manuf 34,* p. 1389–1407, 2021.

[13] G. Álvarez-Narciandi, A. Motroni, M. Pino, A. Buffi und P. Nepa, „A UHF-RFID gate control system based on a Convolutional Neural Network,“ *2019 IEEE International Conference on RFID Technology and Applications (RFID-TA),* pp. 353-356, 2019.

[14] K. Szwedziak, Ż. Grzywacz, E. Polańczyk, P. Bębenek und M. Olejnik, „Optimization of Management Processes in Assessing the Quality of Stored Grain Using Vision Techniques and Artificial Neural Networks,“ *Applied Sciences, Volume 10,* 2020.

[15] D. Tabernik, S. Šela, J. Skvarč und D. Skočaj, „Segmentation-based deep-learning approach for surface-defect detection,“ *Journal of Intelligent Manufacturing,* p. 759–776, 2020.

[16] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes und A. Valenzuela, „A Review of Convolutional Neural Network Applied to Fruit Image Processing,“ *Applied Sciences 10,* 2020.

[17] R. Khan, V. Hemamalini, S. Rajarajeswari, S. Nachiyappan, M. Sambath, T. Devi, B. K. Singh und A. Raghuvanshi, „Food Quality Inspection and Grading Using Efficient Image Segmentation and Machine Learning-Based System,“ *Journal of Food Quality,* 2022.

[18] A. Živković, J. Franjković und D. Dujak, „The role of organizational commitment in employee turnover in logistics activities of food supply chain,“ *Logforum 17,* pp. 25-36, 2021.

[19] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey und I. Sutskever, „Robust Speech Recognition via Large-Scale Weak Supervision,“ *Proceedings of Machine Learning Research,* pp. 28492-28518, 2023.

[20] L. Korel, U. Yorsh, A. S. Behr, N. Kockmann und M. Holeňa, „Text-to-Ontology Mapping via Natural Language Processing with Application to Search for

DOI: 10.2195/lj_proc_ludwig_en_202310_01
URN: urn:nbn:de:0009-14-58111

Relevant Ontologies in Catalysis,“ *Computers 2023, 12(1),* 2022.

[21] Y. Feng, L. Qi und W. Tian, „PhenoBERT: A Combined Deep Learning Method for Automated Recognition of Human Phenotype Ontology,“ *EEE/ACM Transactions on Computational Biology and Bioinformatics,* pp. 1269-1277, 2023.

[22] V. Mijalcheva, A. Davcheva, S. Gramatikov, M. Jovanovik, D. Trajanov und R. Stojanov, „Learning Robust Food Ontology Alignment,“ *2022 IEEE International Conference on Big Data (Big Data),* pp. 4097-4104, 2022.

[23] P. Bojanowski, E. Grave, A. Joulin und T. Mikolov, „Enriching Word Vectors with Subword Information,“ *Transactions of the Association for Computational Linguistics,* pp. 135-146, 2017.

[24] What3words Ltd., „what3words,“ 2023. [Online]. Available: https://what3words.com/about.

[25] F. Cai, W. Zhou , M. Fei und B. Faltings, „Explicit Slot-Intent Mapping with BERT for Joint Multi-Intent Detection and Slot Filling,“ *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 7607-7611, 2021.

[26] M. Honnibal und I. Montani, „spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing,“ 2017. [Online].

[27] M. Honnibal und M. Johnson, „An Improved Non-monotonic Transition System for Dependency Parsing,“ *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing,* pp. 1373-1378, 2015.

[28] C. Shorten, T. M. Khoshgoftaar und B. Furht, „Text Data Augmentation for Deep Learning,“ *Journal of Big Data,* 2021.

[29] A. Schalley, „Ontologies and ontological methods in linguistics,“ *Language and Linguistics Compass, Volume 13,* 2019.

[30] DIN e.V., „DIN ISO 2768-1:1991-06 Allgemeintoleranzen; Toleranzen für Längen- und Winkelmaße ohne einzelne Toleranzeintragung,“ 1991.

Heiner Ludwig, M.Sc., Research Assistant at the Chair of Logistics Engineering, TU Dresden since 2020. Between 2012 and 2020 he studied applied computer science at the TU Dresden.

Address: Technische Universität Dresden, Professur für Technische Logistik, 01062 Dresden, Germany, Phone: +49 351 463-34207, Fax: +49 351 463-35499, E-Mail: heiner.ludwig@tu-dresden.de

Dr.-Ing. Mathias Kühn, Head of Factory Planning Working Group at the Chair of Logistics Engineering, TU Dresden since 2021. Between 2007 and 2014 he studied Mechanical Engineering at the TU Dresden. He received his Ph.D. from the TU Dresden in 2021.

Address: Technische Universität Dresden, Professur für Technische Logistik, 01062 Dresden, Germany, Phone: +49 351 463-32112, Fax: +49 351 463-35499, E-Mail: mathias.kuehn@tu-dresden.de

Prof. Dr.-Ing. Thorsten Schmidt, Head of the Chair of Logistics Engineering, TU Dresden. Prof. Thorsten Schmidt is full professor at the TU Dresden and heads the Chair of Logistics Engineering at the Mechanical Engineering Faculty since 2008. He holds a diploma degree in Mechanical Engineering (TU Dortmund) and a master degree in Industrial Engineering (Georgia Institute of Technology). He received his Ph.D. from the TU Dortmund in 2001. His current research interest include energy-efficient control strategies in material flow, formal verification of control logic, power analysis of distributed and self-controlled systems, lightweight structures in material flow, and stress analysis of wire ropes and timing belts.

Address: Technische Universität Dresden, Professur für Technische Logistik, 01062 Dresden, Germany, Phone: +49 351 463-32538, Fax: +49 351 463-35499, E-Mail: thorsten.schmidt@tu-dresden.de