International Journal of Medical Informatics 160 (2022) 104695

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Combining supervised and unsupervised named entity recognition to detect psychosocial risk factors in occupational health checks

L. Uronen [a,*,1], S. Salanterä [b], K. Hakala [c], J. Hartiala [d], H. Moen [c,1]

[a] Department of Nursing Science, University of Turku, Turku, Finland
[b] Department of Nursing Science, University of Turku, Turku, Finland
[c] Department of Information Technology, University of Turku, Finland
[d] Faculty of Medicine, University of Turku, Turku, Finland

ARTICLE INFO

Keywords:
Health check
Occupational health
Psychosocial risk factors
Text mining, named entity recognition

ABSTRACT

Introduction: In occupational health checks the information about psychosocial risk factors, which influence work ability, is documented in free text. Early detection of psychosocial risk factors helps occupational health care to choose the right and targeted interventions to maintain work capacity. In this study the aim was to evaluate if we can automate the recognition of these psychosocial risk factors in occupational health check electronic records with natural language processing (NLP).

Materials and methods: We compared supervised and unsupervised named entity recognition (NER) to detect psychosocial risk factors from health checks' documentation. Occupational health nurses have done these records.

Results: Both methods found over 60% of psychosocial risk factors from the records. However, the combination of BERT-NER (supervised NER) and QExp (query expansion/paraphrasing) seems to be more suitable. In both methods the most (correct) risk factors were found in the *work environment and equipment* category.

Conclusion: This study showed that it was possible to detect risk factors automatically from free-text documentation of health checks. It is possible to develop a text mining tool to automate the detection of psychosocial risk factors at an early stage.

1. Introduction

Psychosocial risk factors at work influence workability, may cause absence from work or in the worst-case lead to early retirement [1]. Around 10–15% of disability pensions are due to psychosocial work risk factors. Studies have shown that 50% to 60% of all lost working days have links with work-related stress. In the European Union the yearly cost of work-related stress is around EUR 40,000 million each year [1,2]. Physical workloads, computer work, and low job control are risks for disability retirement [3,4]. All psychosocial factors, especially hazardous exposure, job demands, social relationships, workplace violence, and shift work, have a strong association with long sickness absences [1,5]. A European-level collaboration (Prima EF) has developed a psychosocial risk management standard, the PAS 1010 standard, to help detect and maintain people's ability to work. This standard provides guidance, recommendations, and a classification of 13 psychosocial factors that should be recognized in occupational health care. We refer to these are risk factor categories, and they are: *job content, workload and work pace, work schedule, work control, environment and equipment, organizational culture and function, interpersonal relationships at work, role in organization, career development, home-work-interface, violence at workplace, harassment, and bullying* [6,7]. In this study the aim was to automate the recognition of these psychosocial risk factors in occupational health check electronic records that are documented as free text by occupational health nurses.

The basic task of occupational health care is to promote employees' work ability and to prevent disability throughout their careers. The aim is to detect a decrease in an employee's work ability and plan supportive actions in good time. Occupational health checks are one possibility for detecting risk factors of work and workability at an early stage. In occupational health care, these health checks are job specific and designed for the company's goals and exposures [8]. In health checks, an

* Corresponding author at: Department of Nursing Science, 20014, University of Turku, Finland.
E-mail addresses: leenauronen@utu.fi, leena.uronen@mehilainen.fi (L. Uronen).
[1] Equal contributors.

https://doi.org/10.1016/j.ijmedinf.2022.104695
Received 13 August 2021; Received in revised form 25 October 2021; Accepted 16 January 2022
Available online 3 February 2022

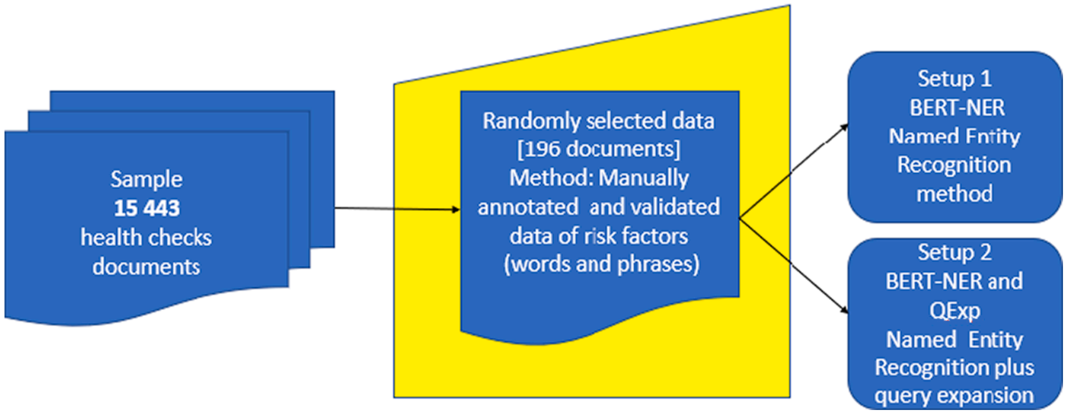
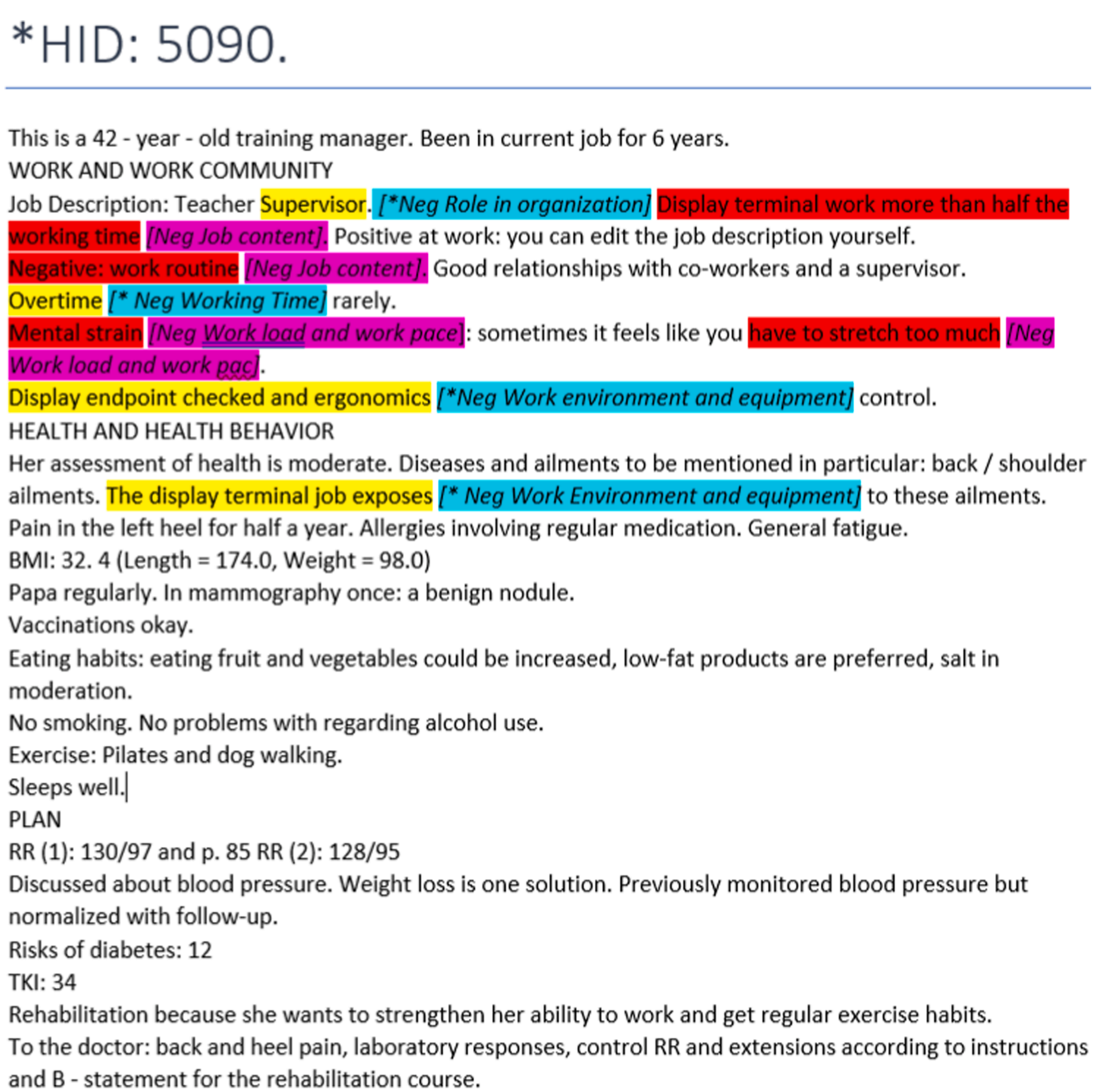


Fig. 1. Workflow of the study.

occupational health nurse (OHN) gathers information about issues that influence work ability or health. However, there is only little research on health checks in occupational care. There is also some research on the effectiveness of these health checks. In review of general health checks the conclusion was that the health checks did not reduce morbidity or mortality [9]. The documents of occupational health checks, typically performed by occupational health nurses, contain valuable information about psychosocial risk factors at work. Compared to structured data,



*HID: 5090.

This is a 42 - year - old training manager. Been in current job for 6 years.
WORK AND WORK COMMUNITY
Job Description: Teacher Supervisor. [*Neg Role in organization] Display terminal work more than half the working time [Neg Job content]. Positive at work: you can edit the job description yourself.
Negative: work routine [Neg Job content]. Good relationships with co-workers and a supervisor.
Overtime [* Neg Working Time] rarely.
Mental strain [Neg Work load and work pace]: sometimes it feels like you have to stretch too much [Neg Work load and work pac].
Display endpoint checked and ergonomics [*Neg Work environment and equipment] control.
HEALTH AND HEALTH BEHAVIOR
Her assessment of health is moderate. Diseases and ailments to be mentioned in particular: back / shoulder ailments. The display terminal job exposes [* Neg Work Environment and equipment] to these ailments.
Pain in the left heel for half a year. Allergies involving regular medication. General fatigue.
BMI: 32. 4 (Length = 174.0, Weight = 98.0)
Papa regularly. In mammography once: a benign nodule.
Vaccinations okay.
Eating habits: eating fruit and vegetables could be increased, low-fat products are preferred, salt in moderation.
No smoking. No problems with regarding alcohol use.
Exercise: Pilates and dog walking.
Sleeps well.
PLAN
RR (1): 130/97 and p. 85 RR (2): 128/95
Discussed about blood pressure. Weight loss is one solution. Previously monitored blood pressure but normalized with follow-up.
Risks of diabetes: 12
TKI: 34
Rehabilitation because she wants to strengthen her ability to work and get regular exercise habits.
To the doctor: back and heel pain, laboratory responses, control RR and extensions according to instructions and B - statement for the rehabilitation course.

Fig. 2. Health check documentations with BERT-NER findings (findings: red, psychosocial risk category: violet) and with QExp findings (findings: yellow, psychosocial risk category; blue) *HID means Hidden IDentification, the number that has been given when the data has been deidentified.

free text can be demanding to analyze, and it is often difficult for nurses to form a comprehensive picture of a person's ability to work based on these documents. This is especially the case when the amount of text is relatively large. Further, free text makes it difficult to use these otherwise valuable and rich records to support health professionals in decisions concerning risks and finding suitable interventions. However, free text gives a possibility to write about the customers story in their own words and describe individual situations that can not to be structured [10,11].

Natural language processing focuses on using artificial intelligence and machine learning to automate the processing of language data, one example being nursing text [12–14]. Named entity recognition (NER) focuses on automated recognition of text spans that refer to certain predefined categories (named entities) [15]. To train a NER model, manual annotation is typically required to first generate the required training data. For specialized domains like medicine and nursing, this includes having domain experts manually labelling a text sample. Here NER is commonly used to identify expressions of clinically significant entities such as events, interventions, diagnoses, diseases, and drugs in free-text narratives [16,17,18,19]. A central challenge in NER is to make a system that can automatically recognize and label words and phrases in unlabeled text that has the same meaning as the categories which the labels represent.

In this paper we describe our work on using NER for automating the task of detecting psychosocial risk factors from the free-text records from occupational health checks. An ensemble system is explored which uses a supervised NER method, trained on a manually annotated dataset, and which is further supplemented with a novel unsupervised paraphrasing method. The paraphrasing method is used to search more loosely for words and phrases that have potentially been missed by the supervised NER method/model, whose linguistic composition may differ quite a bit from the original (manual) annotations, while still having similar meaning. The aim of this paper is to determine if this approach is promising for use in a system that automatically identifies psychosocial risk factors in the free-text reports from occupational health checks.

2. Methodology

2.1. Material

This study is based on a retrospective register of occupational health check records documented by occupational health nurses. The data consists of 15 443 electronic health check records from 7078 employees, recorded in a nationwide private occupational health service in Finland from 2002 to 2007. The data set was filtered by selecting records of employees who were at least 18 years old. This study has ethical approval and permission to use the data from the Ministry of Social Affairs and Health in Finland.

As part of a previous study [11], a random sample of 196 health check reports were manually annotated with psychosocial factors. The annotation scheme and guideline were derived from PAS 1010. This consisted of the following 13 categories: *job content, workload and work pace, work schedule, work control, environment and equipment, organizational culture and function, interpersonal relationships at work, role in organization, career development, home-work-interface, violence at workplace, harassment, and bullying* [7]. The annotation scheme included both positive and negative mentions associated with the different risk factors.

In the present study we were only after detecting the negative mentions - which we here refer to as psychosocial risk factors. The data was pre-processed by using the NLTK word tokenizer for Finnish [20]. The text was also lowercased.

2.2. Methods

Our ensemble system uses a supervised NER method supplemented with an unsupervised paraphrasing method. Named entity recognition (NER) is a basic task in information extraction and has the ability to detect mentions of domain-relevant entities [21]. BERT stands for Bidirectional Encoder Representations from Transformers and is designed to pre-train deep bidirectional representations from unlabelled text [22] (See Fig. 1).

2.2.1. Supervised NER - BERT-NER

The supervised NER method (Setup 1) follows the approach by [21], relying on a pre-trained BERT model [22] and a conditional random field [23] output layer. We refer to this method/model as BERT-NER. The model was pretrained on a general domain Finnish corpus [24] and subsequently fine-tuned on the manually annotated data - the 196 health check reports. When the model was trained, it was used to automatically annotate unseen text data by labelling each word with an "O" label (not a risk factor), or label corresponding to one of the 13 risk factor categories. See Fig. 2 for an example, Bert-NER findings are in red. We split the annotated data into training, development, and test sets. The development set was used for early stopping, but no hyper-parameter adjusting was done for the model. Early stopping is a method that allows to specify an arbitrary large number of training epochs and stop training once the model performance stops improving [22]. The trained NER model achieved an F1 score of 76% on the test set (precision = 75.73%, recall = 76.56%).

2.2.2. Unsupervised NER – BERT-NER and QExp

The unsupervised method is based on the method introduced in [25]. We refer to this method as QExp (Setup 2). Its aim is to search more loosely to find relevant risk factor mentions that have potentially been missed by the BERT-NER method/model. As an input it takes a query, which here is a single risk factor example (word or phrase). It then performs what one may classify as query rewriting or paraphrasing to find words and phrases in the targeted text data whose linguistic composition may differ from the query, while still having meanings that are similar to that expressed by the query.

This method relies on a combination of primarily three components: A semantic word n-gram vector model trained with the word2vec toolkit [26] on a large corpus of clinical text (130 M tokens); A statistical language model trained with the KenLM toolkit [27] on the dataset; And a document search engine in the form of Apache Solr [28]. Briefly explained, the way the method works is by first generating a set of plausible phrases (rewrite) candidates for a given query. This is done by first composing vector representation(s) of the query, and then searching for and retrieving word n-grams that are close by in the semantic vector space. These n-grams are then concatenated to form the phrase candidates. In this process, the statistical language model helps to quickly discard phrases (i.e., word sequences) that are likely nonsensical. Next the phrases are ranked according to their similarity to the query (again using the semantic model), and finally the search engine checks which phrase candidates actually exist in the targeted corpus, and where. Matching phrases are annotated and assigned the same category label as the query. For more information, please see [25]. Due to this method working in an unsupervised manner, unlike BERT-NER, this method does not require any labelled training data to work. This means that we can simply give it lists of example phrases as queries to search for. See Fig. 2 for an example, the entities found by the QExp method are marked in yellow.

To remove some of the obvious mistakes that it makes when used for annotating the text (its false positives), we first ran it on the annotated dataset to form a blacklist of terms (words and phrases) that should not be included. Here we primarily used manual assessment by a domain expert to filter out the incorrect annotations it made. These incorrect annotations, i.e., the incorrect terms found, were used to form a blacklist to filter out these in the experiment presented below (independent of the context in which they occur).

Table 1
Experimental results for the different setups. When calculating Precision, Recall and F1-score, the left side of the bar shows the scores where "Risk factor but wrong category" is considered as false positives, while the right side of the bar shows the scores when these are counted as true positives.

	Setup 1	Setup 2	
	Results	**Additional mentions found**	**Results**
Correct risk factors found (true positives)	296 (59.44%)	55 (11.04%)	**351** (70.48%)
Risk factor found but in wrong category	23 (4.62%)	9 (1.81%)	**32** (6.43%)
Non-risk factors (false positives)	**39** (-)	107 (-)	146 (-)
Missing risk factors (false negatives)	179 (35.94%)	116 (23.29%)	
	Risk factor found but in wrong category counted as: false positives|true positives		
Precision	**0.8268** | **0.8911**	0.6635 | 0.7240	
Recall	0.5944 | 0.6406	**0.7048** | **0.7691**	
F1	**0.6916** | 0.7453	0.6835 | **0.7459**	

2.2.3. *Experiment - application of the methods*

To evaluate our approach, we extracted an evaluation set containing 100 randomly selected health check reports (not overlapping with the 196 originally annotated dataset). We applied our system with the different setups listed below (1 and 2). Two setups of the approach were tested and evaluated:

- Setup 1: Here we applied the supervised BERT-NER model alone, trained on the annotated dataset.
- Setup 2: Here we combined BERT-NER (trained on the annotated dataset) with the QExp method where we give it the originally annotated phrases (i.e., the queries are the text spans that has been assigned a risk factor label by the human annotators).

Both setups (100 reports each) were evaluated manually by a domain expert based on documentation (Fig. 2) what was found in terms of:

- Correct risk factors found (true positives).
- Risk factors found that have the wrong risk factor category/label.
- Non-risk factors (false positives).
- Missing risk factors (false negatives).

3. Results

In our work we used NER for automating the task of detecting psychosocial risk factors from the free-text documentation from occupational health checks. We combined supervised NER method (BERT-NER), trained on a manually annotated dataset with a novel unsupervised paraphrasing method (QExp). The results are shown in Table 1.

In Setup 1 we evaluated the supervised BERT-NER model alone, trained on the annotated dataset. With Setup 1, 296 correct psychosocial risk factors were found. 23 of the risk factors were labelled with the wrong category. For example, the phrase "can't stand this job" was labelled with the *career development* category and it should have been in the category of *job content*. Altogether 39 non-relevant psychosocial risk factors were found. This setup did not recognize 179 risk factors.

In Setup 2 we combined BERT-NER (trained on the annotated dataset) with the QExp method where we gave it the originally annotated phrases as search queries (i.e., the text spans that have been assigned a risk factor label by the human annotators in the annotated dataset). With Setup 2, a total of 351 correctly annotated risk factors were found, 31 were assigned to the wrong category, and 145 non-relevant psychosocial

Table 2
Confusion matrix in setup 1.

Setup 1	Predicted label												
True label	Categories of psychosocial risk factors	O	Control	Organizational culture and function	Role in organization	Job content	Interpersonal relationships at work	Work schedule	Work load and work pace	Environment and equipment	Career development	Violence at workplace	Home-work-interface
	O	0	6	0	1	8	0	6	6	5	0	0	7
	Control	13	11	1	0	0	0	0	1	0	2	0	0
	Organizational culture and function	4	0	10	0	0	0	0	0	0	0	0	1
	Role in organization	23	0	0	0	0	0	0	0	0	0	0	0
	Job content	16	0	0	1	40	0	1	2	2	0	0	2
	Interpersonal relationships at work	4	0	0	0	0	6	0	0	0	0	0	0
	Work schedule	5	0	0	0	0	0	14	0	0	0	0	0
	Work load and work pace	23	0	0	0	0	0	4	65	0	1	0	0
	Environment and equipment	61	0	0	0	0	0	0	0	106	0	0	0
	Career development	6	1	0	0	3	0	0	0	0	6	0	0
	Violence at workplace	3	0	0	0	0	0	0	0	0	0	0	0
	Home-work-interface	21	0	0	0	0	0	0	1	0	0	0	38

Table 3
Confusion matrix in setup 2.

Setup 2	**Predicted label**												
True label	**Categories of psychosocial risk facto**	**O**	**Control**	**Organizational culture and function**	**Role in organization**	**Job content**	**Interpersonal relationships at work**	**Work schedule**	**Work load and work pace**	**Environment and equipment**	**Career development**	**Violence at workplace**	**Home-work-interface**
	O	0	6	1	10	10	13	17	13	37	0	0	35
	Control	13	11	1	0	2	0	0	2	0	2	0	0
	Organizational culture and function	4	0	10	3	0	0	0	0	0	0	0	1
	Role in organization	13	0	0	6	0	0	0	0	0	0	0	0
	Job content	9	0	0	2	45	0	1	2	2	0	0	2
	Interpersonal relationships at work	3	0	0	0	0	7	0	0	0	0	0	0
	Work schedule	4	0	0	0	0	0	15	0	0	0	0	0
	Work load and work pace	16	0	0	0	0	0	4	71	1	1	0	0
	Environment and equipment	36	0	0	0	0	0	0	0	129	0	0	0
	Career development	4	1	0	0	3	0	0	0	0	8	0	0
	Violence at workplace	3	0	0	0	0	0	0	0	1	0	0	0
	Home-work-interface	9	0	0	0	0	0	0	1	0	0	0	50

risk factors were found. This means that the QExp found 55 new risk factors in correct categories (351 in total) and 8 that were labelled with the wrong category (31 in total). It also incorrectly labelled 106 psychosocial risk factor mentions that were not relevant, which sums up to 145. This setup did not recognize 116 risk factors, so called missing factors. Most often the risk factors that were not found were in the *home-work-interface* category. A typical example of a non-relevant factor was "travel work: no" in which the "travel work" is correctly a risk factor but not in this case when there is the word no attached.

In both methods by far the most (correct) risk factors were found in the *work environment and equipment* category. This category includes factors such as "inadequate equipment availability, suitability or maintenance; poor environmental conditions such as lack of space, poor lighting, excessive noise". The second most occurring risk factors were in the *workload and work pace* category. In this category there are "work overload or under load, machine pacing, high levels of time pressure, continually subject to deadlines". The third category of the most occurring risk factors was *job content*. These include risk factors such as "lack of variety or short work cycles, fragmented or meaningless work, under use of skills, high uncertainty, continuous exposure to people through work". There were no mentions found in the category of bullying. The category of risk factor found but in the wrong category were totally 32 matches and the categories are described in the confusion matrix on Table 2 and 3.

In parallel evaluation the second researcher read 10 documents. These documents had 165 annotations. One mistake was found from all the documents, which were in all three methods. The error rate was 2,4% of all 165 entries.

4. Discussion

The aim of this paper was to explore the use of named entity recognition to detect psychosocial risk factors in occupational health checks and to determine if such an approach is promising for use in a system that automatically identifies psychosocial risk factors in the free-text reports from occupational health checks. Two setups were tested. One that relies on a more traditional NER method trained in a supervised manner to mimic the manually annotated data. In the second setup, the former method is supplemented by an unsupervised paraphrasing method that aims to find potentially additional risk factors.

The combination of BERT-NER (supervised NER) and QExp (query expansion/paraphrasing) seems to constitute the more suitable system for recognizing psychosocial risk factors from free text health checks' documentation when the two methods were compared. The system represented by Setup 2 correctly identified 70% of the psychosocial risk factors from health check records. If correctly identified but classified in the wrong category is included, the result is 77%. We see these results as promising, and they support the earlier efforts towards automatic recognition and labelling of words and phrases in unlabelled clinical text that have the same meaning as the categories which the labels represent [17,19].

The F1 score was calculated as the harmonic mean of precision and recall. However, given that the purpose of this system is to detect and highlight possible risk factors, in a realistic use scenario one could argue that recall is more important than precision. The main reason for this argument is that it is easier for the user to discard non-relevant risk factors than to find the relevant ones. In other words, detecting a larger number of risk factors, even if that means detecting more non-relevant risk factors, is more important than detecting fewer non-relevant ones at the expense of detecting fewer relevant ones.

There are some limitations in this study. One limitation of the study is that there is no gold standard of documentation of the psychosocial risk factors. However, we used PAS 1010 as the annotation scheme, based on a comprehensive review of psychosocial risk management models across Europe. There might be other psychosocial risk factors that we did not evaluate. However, the models used in this study allow new words and phrases to be added. Another limitation is that the data were produced by occupational nurses only from one organisation and thus they might consist of slang specific only to this organisation. However, the organisation operates in several locations nationwide. In addition, the documents were manually annotated by one domain expert and only 10% of the results were double checked by two experts. There is a limited amount of research concerning risk factors in occupational health care and we choose those that were classified in PAS1010.

5. Conclusion

It is possible to automate the detecting of psychosocial risk factors from free text. Both setups found more than 60% of the psychosocial risk factors at work. With additional research and work, it is possible to build a system or tool that facilitates the work and decision-making of an occupational health nurse. Such a tool would highlight and warn the occupational health nurse on possible risk factors during the process of writing the health checks, or retrospectively. With further development, the tool could potentially find information on all previous records and thus help maintain work capacity and promote the work ability literacy of occupational health professionals. This could also support research into ways for how to improve work satisfaction in the population.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Slany, S. Schütte, J.-F. Chastang, A. Parent-Thirion, G. Vermeylen, I. Niedhammer, Psychosocial work factor and long sickness absence in Europe, Int. J. Occ. Env. Health. 20 (1) (2014) 16–25.
[2] P. Suadicani, J.P. Bonde, K. Olesen, F. Gyntelberg, Job satisfaction and intention to quit the job, Occ. Med. 63 (2) (2013) 96–102.
[3] K. Christensen, H. Fevile, M. Labriola, T. Lund, The impact of psychosocial work environment factors on the risk of disability pension in Denmark, Eur. J. Public Health 18 (3) (2008) 235–237.
[4] E. Lahelma, M. Laaksonen, T. Lallukka, P. Martikainen, O. Pietiläinen, P. Saastamoinen, R. Gould, O. Rahkonen, Working conditions as risk factors for disability retirement: a longitudinal register linkage study, BMC Public Health 12 (1) (2012), https://doi.org/10.1186/1471-2458-12-309.
[5] M. Laaksonen, J. Pitkäniemi, O. Rahkonen, E. Lahelma, Work arrangements, physical working conditions, and psychosocial working conditions as risk factors for sickness absence: Bayesian analysis of prospective data, Ann. Epidemiol. 20 (5) (2010) 332–338.
[6] S. Leka, E. Kortum, A European framework to address psychosocial hazards, J. Occ. Health 50 (3) (2008) 294–296.
[7] S. Leka, T. Cox, Guide prima-EF. Guidance on the European framework for psychosocial risk management: A resource for employers and workers representatives. Place of publication: WHO; 2011 [cited 2020 May]. Available from: http://www.prima-ef.org/uploads/1/1/0/2/11022736/prima-ef_brochure_english.pdf.
[8] T. Leino, M. Rautio, M. Kanervisto, J. Tilli, S. Kaleva, Medical examination practices in Finnish occupational health services. Research report. Place of publication: Työterveyslaitos: 2014. [cited 2020 May]. Available from: https://www.julkari.fi/bitstream/handle/10024/120383/Terveystarkastuskaytannot_suomalaisessa_tyoterveyshuollossa.pdf?sequence=1&isAllowed=y.
[9] L.T. Krogsbøll, K.J. Jørgensen, C.G. Larsen, et al., General health checks in adults for reducing morbidity and mortality from disease, Cochrane Database Syst. Rev. (2012:10.). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009009.pub2/full.
[10] W. Blair, B. Smith, Nursing documentation: frameworks and barriers, Contemporary Nurse. 41 (2) (2012) 160–168.
[11] L. Uronen, J. Heimonen, P. Puukka, K.-P. Martimo, J. Hartiala, S. Salanterä, Health check documentation of psychosocial factors using the WAI, Occ. Med. 67 (2) (2016) 151–154.
[12] H. Sookyung, B.J. Stephen, S. Bakken, Exploring the Ability of Natural Language Processing to Extract Data from Nursing Narratives, Comput. Inform. Nurs. 27 (4) (2009) 215–225.
[13] S. Velupillai, D. Mowery, B.R. South, et al., Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis, Yearb. Med. Inform. 10 (1) (2015) 183–193.
[14] D. Albright, A. Lanfranchi, A. Fredriksen, W.F. Styler, C. Warner, J.D. Hwang, J. D. Choi, D. Dligach, R.D. Nielsen, J. Martin, W. Ward, M. Palmer, G.K. Savova,

Towards comprehensive syntactic and semantic annotations of the clinical narrative, J. Am. Med. Inform. Assoc. 20 (5) (2013) 922–930.

[15] D. Jurafsky, J.H. Martin, Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition, Prentice Hall, 2008 https://www.pearsonhighered.com/assets/samplechapter/0/1/3/1/0131873210.pdf Practise Hall, 2008.

[16] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, J.F. Hurdle, Extracting information from textual documents in the electronic health record: A review of recent research, Yearb Med. Inform. 17 (01) (2008) 128–144.

[17] M. Skeppstedt, M. Kvist, G. Nilsson, H. Dalianis, Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study, J. Biomed. Inform. 49 (2014) 148–158.

[18] H. Moen, K. Hakala, F. Mehryary, L.M. Peltonen, T. Salakoski, F. Ginter, S. Salanterä, Detecting mentions of pain and acute confusion in Finnish clinical text, in: BioNLP 2017, 2017, pp. 365–372.

[19] M. Topaz, L. Murga, K.M. Gaddis, M.V. McDonald, O. Bar-Bachar, Y. Goldberg, K. H. Bowles, Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches, J. Biomed. Inform. 90 (2019) 103103, https://doi.org/10.1016/j.jbi.2019.103103.

[20] S. Bird, E. Klein, E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, O'Reilly Media, Inc., 2009.

[21] K. Hakala, S. Pyysalo, Biomedical named entity recognition with multilingual BERT. Proceedings of The 5th Workshop on BioNLP Open Shared Tasks, 2019.

[22] Jacob Devlin, et al., Bert: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT, 2019.

[23] Lafferty, John, Andrew McCallum, Fernando CN Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, 2001.

[24] Virtanen, Antti, et al. Multilingual is not enough: BERT for Finnish, arXiv preprint arXiv:1912.07076, 2019.

[25] H. Moen, L.M. Peltonen, H. Suhonen, H.M. Matinolli, R. Mieronkoski, K. Telen, S. Salanterä, An Unsupervised Query Rewriting Approach Using N-gram Co-occurrence Statistics to Find Similar Phrases in Large Text Corpora, in: Proceedings of the 22nd Nordic Conference on Computational Linguistics, 2019, pp. 131–139.

[26] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, vol. 26, 2013, pp. 3111–3119.

[27] K. Heafield, KenLM: Faster and smaller language model queries, in: Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 187–197.

[28] Apache Solr search platform, https://solr.apache.org (last accessed 22.07.2021).