

Supporting Care by Interpretation of Expressions about Patient Experience with Machine Learning

Lauri Lahti

Abstract— Our research aims at addressing the needs of developing data analysis about communication in respect to care seeking and primary care, discovering how health expressions evolve along the personal growth and learning process, and how to solve the needs identified in respect to developing measuring the quality of life. We provide an overview of the development of a new research methodology exploiting machine learning for analyzing patient experience expressions to support personalized care and managing in everyday life. Our research relies on an online questionnaire in which the representatives of various population groups perform interpretation tasks. Dependencies between answers about the interpretation tasks and background information are analyzed with machine learning methods. The research creates new ways to interpret and address the meanings of language usage of different groups of patients and impaired carefully and distinctively as a part of everyday life and care events.

Index Terms—machine learning, patient experience, personalized healthcare, semantic analysis.

I. INTRODUCTION

We provide an overview of the development of a new research methodology exploiting machine learning for analyzing patient experience expressions to support personalized care and managing in everyday life. Our description relies on the detailed planning and implementation we have made for a new research project titled “Development of method for interpretation of health expressions based on machine learning to support various care events and persons (DIHEML research project)”. The research idea and the research plan for DIHEML research project have been conceived and developed by Lauri Lahti and they have been published in an initial form and concerning main features in his scientific publication (Lahti 2017, <http://urn.fi/URN:NBN:fi:aalto-201712298340>). Lauri Lahti is the responsible researcher of DIHEML research project.

DIHEML research aims at developing a method based on machine learning (artificial intelligence) that enables semantic classification of health expressions and dialogues concerning various events, processes and persons in healthcare. To develop computational methods DIHEML research needs to acquire a broad data collection among the groups of patients and impaired and other population groups about essential viewpoints concerning health and wellbeing. DIHEML research gathers data with an online questionnaire that asks a person to interpret health expressions by giving

answers on various measurement scales. In addition, the person is asked to give answers about his/her background. In brief, the research aims at finding out what kind of dependencies there emerge between the answers about the interpretation tasks and background information. A successful implementation of DIHEML research relies on a diverse set of collaborating organizations that enable a fertile acquisition of data from various fields of life and a detailed analysis with multidisciplinary research partners.

II. PREVIOUS RESEARCH

Previous research has identified the importance of developing computational methods that enable interpretation of biomedical measurement data with a natural language used in thinking and communication (Califf 2014; Tsai et al. 2015). Previous research has also explored ways to identify significant patterns in linguistic health data (Brown et al. 2010; Ashutosh 2014) and how different populations interpret linguistic expressions in respect to affectivity (Bradley & Lang 1999a; Warriner et al. 2013) and associations (Higginbotham et al. 2015; Fitzpatrick et al. 2015).

It has been suggested that various characteristics of a person have an impact on the state of the health (Marmot et al. 2003). An aspect of a patient’s health status that can be directly measured by the patient himself/herself (i.e. without a need for an interpretation of the patient’s response by someone else) can be referred to as a patient-reported outcome and it can be typically measured with a self-administered questionnaire referred to as a patient-reported outcome measure (PROM) (Prinsen et al. 2018). Specific guidelines have been developed to enable systematic reviews of PROMs addressing among other things reliability, validity and sensitivity to change (de Vet et al. 2011).

Health-related quality of life can be measured with both generic and disease-specific instruments (Solans et al. 2008). Generic instruments typically have a more general formulation thus enabling comparison of different patient groups whereas disease-specific instruments typically enable more detailed evaluation of group-specific properties. Aalto et al. (2013) provide a review of generic quality of life measures that have been used in health research carried out with large samples of Finnish population, including 15D (Sintonen 1994), EQ-5D (EuroQol Group 1990), RAND-36 (Hays et al. 1993), WHOQOL-BREF (WHO 1997) and EuroHIS-8 (Nosikov & Gudex 2003).

Information concerning how to support health and wellbeing is being increasingly available online for anyone,

Lauri Lahti, Department of Computer Science, Aalto University School of Science, Finland.

including also descriptions about diagnosis and treatment. This online information can offer resources for developing computational methods to support care with a wide range of information content, such as authorized healthcare guidelines (Terveyskirjasto 2018; Lahti 2016) and general discussion forums (The Suomi 24 Corpus 2016; Lahti et al. 2018).

In our previous work (Lahti 2017) we trained a convolutional neural network model (adapted from the model of Kim (2014)) with sentences representing two classes: 1000 sentences from the discussion topic group Children's health (Lasten terveys, in Finnish) and 1000 sentences from the discussion topic group Health (Terveys, in Finnish) based on Finnish online discussions (The Suomi 24 Corpus 2016). Then the trained model managed to classify correctly 88,6 percent of the new Children's health sentences and 91,4 percent of the new Health sentences.

III. METHOD

Motivated by the previous research DIHEML research aims at developing a method based on machine learning that enables semantic classification of health expressions and dialogues concerning various events, processes and persons in healthcare. Our research aims at addressing the needs of developing data analysis about communication that can support evaluation of the urgency of the need for treatment and decision making in respect to care seeking and primary care. In addition, our research aims at discovering how the interpretations of health expressions evolve along the personal growth and learning process and furthermore among students who are studying to become healthcare professionals and how this can enable supporting the educational system and learning methods. In addition, our research aims at addressing the needs identified in respect to developing measuring the quality of life.

A. Acquisition of data

The acquisition of data for DIHEML research relies on an online questionnaire in which the person participating in the research is asked to give an answer on various measurement scales to a) claims shown in text, b) images and c) videos. These answers indicate how the person interprets the shown health expressions and this activity is referred to as performing *health expression interpretation tasks*. In addition, in the acquisition of data the person is asked to give answers about his/her background concerning the conditions of life and quality of life, including information of the state of health and the contact information to enable later longitudinal research.

The answers about interpretation tasks and background which are gathered from the person can be considered as a collection of patient-reported outcome measures (PROMs) that are used to identify patterns of health behavior. The research aims at finding out what kind of dependencies there emerge between the answers about the interpretation tasks and background information given by the person participating in the research. Based on the gathered data the research develops machine learning methods (artificial intelligence) that support care and wellbeing. The gathered data and methods developed based on its analysis can be used in various ways to implement different personalized health

services and to advance addressing different needs in everyday life and care. DIHEML research is carried out in the context of Finnish language usage and Finnish healthcare system and society but the results are expected to be generalized to other contexts as well when taking into account the similarities and differences in respect to the Finnish context.

The population groups asked to answer to the online questionnaire consist of a) the groups of patients and impaired, and b) the students of primary school, secondary school, high school and educational institutes in the domain of healthcare. The groups of patients and impaired are recruited via organizations of patients and impaired, and the students are recruited via the authorities of educational institutes. A goal is to gain 5000 persons participating in the research. The acquisition of data is based on various different experiment series (online sessions). The person can choose himself/herself how many experiment series he/she wants to participate. An assumption is that one experiment series lasts about one hour.

B. Formulation of interpretation tasks based on health expressions

In brief, DIHEML research relies on a collection of thousands of health expressions and with randomization a set of 100 expressions at a time are given to the person to be interpreted by giving answers on various measurement scales. In addition, the person is asked to give answers about his/her background.

In DIHEML research the patient experience is considered to modularly consist of *health expression events* (HEE) and a health expression event can be described with a *health expression* (HE), for example "I have a feeling of illness" or "I am in a situation where the doctor tells me the diagnosis".

A *health expression event series* (HEES) is a group of health expression events that occur *as a consecutive series* and/or *as a parallel series*. In practice a health expression event series can be described as a kind of *dialogue* which possibly contains consecutive and/or parallel expressions.

In the questionnaire the interpretation tasks can have various formulations. In the interpretation task the person can be asked to interpret the shown expression like it was an imagined experience of his/her own person (*imagined my person's experience*, see Figures 1a and 1c) or like it was an imagined experience of another person (*imagined other person's experience*, see Figure 1b).

In *rating interpretation* for example the expression "I have a feeling of illness" is shown and then the person is asked to interpret what kind of impression this expression induces in him/her in respect to worriedness, on a scale 0-10 where 0 means the smallest possible worriedness and 10 means the greatest possible worriedness (see Figure 1a). In *comparison interpretation* for example two expressions "I have a feeling of illness" and "I have a feeling of faintness" are shown and then the person is asked to interpret which one of these two expressions induces in him/her a greater impression in respect to worriedness (see Figure 1b).

If the person is asked to interpret for example two expressions that occur as a consecutive series then the interpretation task can be referred to as the *interpretation of consecutive expressions* (see Figure 1c). If the person is asked

to interpret for example two expressions that occur as a parallel series then the interpretation task can be referred to as the *interpretation of parallel expressions*.

An interpretation task can ask the person to give an answer on various measurement scales to claims shown in text, images or videos. Figure 1 illustrates two *text-based interpretation tasks* (a and b) and an *image-based interpretation task* (c).

An image-based interpretation task is created on the basis of a corresponding text-based interpretation task using HEE image templates (template variations currently include “a situation with one human figure where the person thinks” and “a situation with two human figures where the person thinks”; see images 1 and 2 in Figure 1c). In the HEE image templates the simplified form of the human figure, referred to as a tadpole, aims at making the human figure intuitively recognizable for people at any age (Foley & Mullis 2008).

A *video-based interpretation task* is created on the basis of a corresponding text-based interpretation task or image-based interpretation task. Thus a video-based interpretation task is a dynamic slide show which presents consecutively one textual view or image at a time all the stimuli that are included in the corresponding text-based or image-based interpretation task. Video-based interpretation tasks currently include the following variations: transitions automatically, transitions activated by the person, interpretation holistically, interpretation sequentially, interpretation cumulatively, a replay option is offered, and a replay option is not offered.

C. Adapting to the person’s age and sensitivity to become anxious

DIHEML research is active from June 2018 to May 2021. The research is funded by scholarships given for research purposes. The persons participating in the research are not paid for the participation. The research is implemented so that respect is given towards the privacy of the participating persons, the academic research ethical principles, the data protection and privacy regulation of European Union and the

guidelines of the Finnish research ethics authority TENK.

For the acquisition of data in DIHEML research various procedures have been prepared to ensure carefully that the persons participating in the research cannot be harmed in any way. Preparations have been made concerning possible problem situations encountered during the acquisition of data so that when giving answers with the online questionnaire there are guidelines and personal guidance available via a specific online form.

In respect to psychological stress intensity there is an aim to adapt the questionnaire contents (interpretation tasks and questions) to the person’s age and sensitivity to become anxious which are asked instantly in the beginning of the data acquisition. This adaptation is carried out according to the corresponding age limits defined for media contents (KAVI 2018), and following them Lauri Lahti has classified all information that is shown in the DIHEML research.

In the following are some examples of expressions belonging to each psychological stress intensity class: “I have pain.” (permitted to people of any age), “I have a bleeding nose.” (at least 7-year-olds), “I have an open fracture.” (at least 12-year-olds), “I have anorexia eating disorder.” (at least 16-year-olds), and “I have suicidal thoughts.” (at least 18-year-olds).

In the evaluation of the sensitivity to become anxious the person is asked to answer to the part “generalized anxiety” in the SCARED questionnaire (Birmaher et al. 1997; Birmaher et al. 1999; Hale et al. 2005; Bögels & Van Melick 2004). If the total score for the person’s answers is below the threshold value 9 then the research will show to the person contents that belong to the intensity class of his/her age or belong to lower intensity classes (thus the sensitivity to become anxious has been considered low for the person). If the total score for the person’s answers is at least the threshold value 9 then the research will show to the person contents that belong only to the intensity class “permitted to people of any age” (thus the sensitivity to become anxious has been considered high for the person).

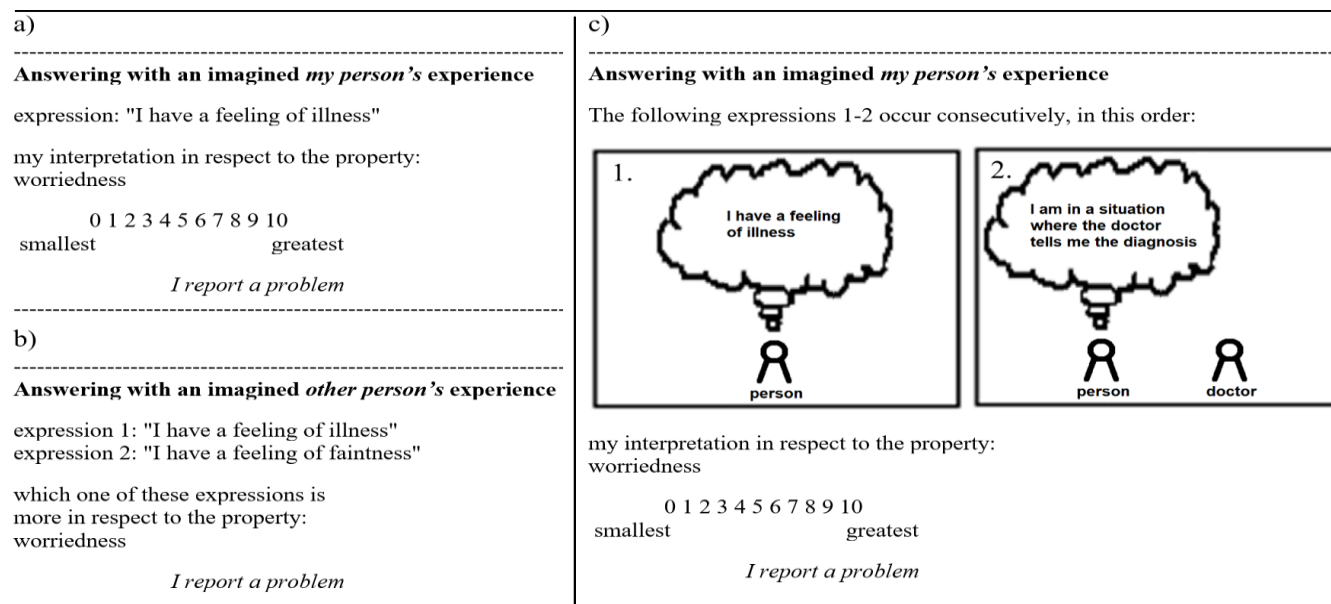


Figure 1. Illustrations about two text-based interpretation tasks (a and b) and an image-based interpretation task (c).

D. Authorization of the caregiver

The information document about DIHEML research which contains an online link to the questionnaire is given to the persons younger than 15-year-olds only after it has been confirmed that an authorization has been gained from the caregiver named for the person in question. This confirmation is performed by the responsible researcher of DIHEML research or someone who assists him (for example a teacher of an educational institute). The confirmation is carried out so that an information document is sent to the caregiver of the person participating in the research in such a way that this information document contains an online link to a form that enables giving the authorization.

When the caregiver is giving the authorization, the caregiver expresses the age and sensitivity to become anxious concerning the cared person. This information sets the initial upper limit for the psychological stress intensity class (the age and sensitivity to become anxious expressed by the cared person himself/herself can set the upper limit for the intensity class still even lower).

In addition, the caregiver is required to give beforehand an answer whether or not the cared person is allowed to start answering the questionnaire independently, or is the cared person allowed to start answering the questionnaire only when the caregiver is present or only when the teacher of the cared person is present. Furthermore, the caregiver gets beforehand an access to familiarize himself/herself with and accept the questionnaire contents that are used as a basis for the interpretation tasks of the cared person, and the caregiver is provided with an opportunity to remove tasks that he/she considers as harmful for the cared person so that the cared person will not be exposed to them at all.

IV. EXPERIMENT

As a part of the planning and implementation of DIHEML research project we have carried out an experiment to identify suitable materials and methods to be used in the questionnaire for data acquisition with the representatives of target populations. Especially we have tried to identify suitable formulations concerning the information asked about the background of the person and the interpretation tasks given to the person.

Based on a literature review and some preliminary testing with the representatives of the target populations we suggest that the following entities are useful information to be asked about the background of the person:

- the conditions of life (among others gender, birthtime, place of living, native language, profession, state of health, chronic diseases/disability/impairment, gained care for chronic diseases/disability/impairment)
- responses about the age and sensitivity to become anxious answered by the person and by the caregiver (Birmaher et al. 1997; Birmaher et al. 1999; Hale et al. 2005; Bögels & Van Melick 2004; Spitzer et al. 2006)
- measures of quality of life (including 15D (Sintonen 1994), EQ-5D (EuroQol Group 1990), EQ-5D-5L (Herdman et al. 2011), RAND-36 (Hays et al. 1993), WHOQOL-BREF (WHO 1997) and EuroHIS-8 (Nosikov & Gudex 2003)).

Based on a literature review and data analysis that Lauri

Lahti carried out to extract health expressions from various information resources we suggest that the following entities are useful interpretation tasks to be given to the person:

- health expression interpretation tasks in text, image and video format concerning everyday life based on the resource Lahti (2018a, "Expression collection of everyday life")
- health expression interpretation tasks based on the resources of Bradley and Lang (1999a), Warriner et al. (2013), Lang et al. (2008), Söderholm et al. (2013), and Eilola and Havelka (2010), especially the stimuli representing the extremes of affectivity and health-related topics.

Health expression interpretation tasks rely on the resource Lahti (2018a, "Expression collection of everyday life") that has been conceived and developed by Lauri Lahti. Since this resource lists health expression interpretation tasks, it thus lists also both the expressions that are asked to be measured (measuring material, for example "I have a feeling of illness") and the measuring scales (measuring dimensions, for example "worriedness"). All these measuring material and measuring dimensions are based on text strings that have been considered significant when extracted by Lauri Lahti with algorithms (some algorithms discussed in Lahti et al. (2018); Lahti (2017); Lahti (2016); Finnish Dependency Parser (2018); Dunning (1993)) from various resources, including:

- authorized healthcare guidelines (among others Terveyskirjasto 2018; ICD-10 2011; ICF 2013)
- support materials of the organizations of patients and impaired and other health organizations
- texts of online discussion forums (both general and health-specific forums)
- online publications providing health-related articles
- Lauri Lahti's notes from discussions with the representatives of organizations of patients and impaired, National Institute for Health and Welfare in Finland (THL), and the experts of primary care
- materials and methods motivated by the previous resembling research work (among others Aalto et al. (2013) and Koskinen et al. (2012)).

Two samples from the resource Lahti (2018a, "Expression collection of everyday life") are illustrated in Table 1 and Table 2. Table 1 shows all 10 main categories of health expressions included in the measuring material and the number of expressions and subcategories. Table 2 shows all 52 measuring dimensions (measurement scales used to give answers about interpretations of health expressions).

Resembling our previous work (Lahti 2017) we carried out a new *sentence classification experiment*. We trained a convolutional neural network model (adapted from Kim (2014) and Britz (2015)) with a sample of 2000 unique sentences of Finnish online discussions from the discussion topic group "Terveys" (Health, in Finnish (The Suomi 24 Corpus 2016)). We trained the model with two classes of 1000 sentences: we had labeled beforehand 1000 sentences as pleasurable and 1000 sentences as unpleasurable based on affective ratings of the words occurring in a sentence (Warriner et al. 2013; Bradley & Lang 1999; Söderholm et al.

Table 1. All 10 main categories of health expressions included in the measuring material (Lahti 2018a, "Expression collection of everyday life") and the number of expressions and subcategories.

<i>Name of main category of health expressions</i>	<i>Number of expressions</i>
1. Symptom descriptions (feeling)	290 (13 subcategories)
2. Care event (carried out by healthcare personnel)	263 (4 subcategories)
3. Self-care (carried out by a layman)	97 (1 subcategory)
4. Accident event or exposure to other sad event	420 (12 subcategories)
5. Survival from an accident and readiness to help	70 (1 subcategory)
6. Care seeking, asking or seeking health information	557 (3 subcategories)
7. Chronic disease or impairment and challenges related to it	116 (1 subcategory)
8. Risky event (danger of evolving to an accident)	14 (1 subcategory)
9. Preparedness for an accident	25 (1 subcategory)
10. General words and sentences of everyday life related to care and wellbeing	272 (1 subcategory)

Table 2. All 52 measuring dimensions (measurement scales used to give answers about interpretations of health expressions) (Lahti 2018a, "Expression collection of everyday life").

<i>Name of measuring dimension</i>	<i>Name of measuring dimension (continued)</i>
1. need for help	27. worriedness
2. urgency	28. giving up vs. endurance to fight
3. pain	29. stressfulness
4. scariness	30. dirtiness
5. hopefulness	31. expressiveness concerning empathy
6. health	32. need to get empathy
7. complexness	33. concreteness vs. abstractness
8. sorrow	34. amount of risks
9. pleasure	35. need to search information from the internet
10. arousal	36. need to ask an opinion from at least two doctors about the thing in question
11. dominance	37. imageability of the thing in question
12. shamefulness	38. ability to perform decision making concerning the thing in question
13. emotionality	39. ability to perform procedures concerning the thing in question
14. familiarity	40. desire of seeking care immediately at the emergency room of a healthcare station (due to the thing in question)
15. need to learn about the thing in question	41. desire of seeking care immediately on the next day at the emergency room of a healthcare station (due to the thing in question)
16. need to cry for help	42. desire of booking a non-urgent appointment at a healthcare station (due to the thing in question; an assumption is that thus the admittance to the doctor's appointment is after about a week)
17. need to call to the emergency number 112	43. need to get help from a specialized doctor
18. need to get help from a doctor	44. need to get help from a close family member
19. need to get help from a friend	45. need to get peer-support from persons who experience the same thing
20. ability to independent problem solving	46. ability to a sufficient self-care without a need of seeking care at the emergency room of a healthcare station
21. belief in possibility to survive	47. ability to a sufficient self-care without a need of calling to the emergency number 112
22. ability to be assisting in the thing in question	48. ability to a sufficient self-care without the arrival of a mobile primary care unit to my location
23. how important part of the quality of life	49. carefulness and slowness vs. coarseness and fastness
24. disappointedness	50. how well I understand the thing described by this interpretation task
25. excitedness	51. my expertise about the thing in question
26. anxiety	52. importance of the thing in question when aiming at developing addressing the needs of a patient

2013; Eilola & Havelka 2010). We evaluated the trained model with additional 2000 sentences that were not used in the training. The trained convolutional neural network model managed to classify correctly 79,8 percent of the new 1000 pleasurable Health sentences and 97,2 percent of the new 1000 unpleasurable Health sentences. Details about our sentence classification experiment are available in a supplement to this research article (Lahti 2018b).

V. DISCUSSION

The acquisition of data for DIHEML research project produces a data collection that enables to investigate what kind of dependencies there emerge between the answers about the interpretation tasks and background information given by the person participating in the research. Based on the gathered data the research develops machine learning methods (artificial intelligence) that support the care and wellbeing of the groups of patients and impaired and other population groups. The promising results of our new sentence classification experiment relying on a convolutional neural network model gives support for our central aim to exploit machine learning to identify dependencies in a resembling health-related data that can be gathered in DIHEML research.

All the data gathered in the acquisition of data for DIHEML research is planned to be permanently archived by the research group. In addition information planned to be permanently archived by the research group include the research registry which contains approval documents and documents for decoding code names, and the detailed descriptions about the predictive models and the measurement tool developed in the research as well as the data base consisting of the measurement results. In all this archive data unique personal information is stored, i.e. identification information, which enables later longitudinal research related to the data collection. The processing and storing of personal information are needed to enable combining the person's answers given in separate sessions especially for longitudinal research purposes and also to investigate the dependencies between the person's state of health and his/her other answers.

Besides developing methods the research will publish an appropriately anonymized version of the measurement results gathered by the research freely available for everybody.

DIHEML research creates new ways to interpret and address the meanings of language usage of different groups of patients and impaired carefully and distinctively as a part of everyday life and care events (reflecting the meanings of language usage in respect to among other things context, background and personal profile). The research offers new ways to interpret the language usage of the representatives of groups of patients and impaired so that misunderstanding can be prevented and agreement can be advanced. The research offers also ways to highlight in language usage of the representatives of groups of patients and impaired such topics that are important to be addressed in implementation and development of healthcare services. Thus the research aims at making such a progress that the personal needs and rights of patients and impaired can be better addressed in decision making. The research aims at advancing creation of support

services needed by different special population groups in everyday life and care. Thus the research enables advancing the equality for different population groups and for population groups that have special needs.

The answers gained from the students of educational institutions supplement the answers gained from the groups of patients and impaired, thus also providing comparison data representing the general population. In addition an opportunity is achieved for finding out how the interpretations evolve along the personal growth and learning process and furthermore separately among the persons attending educational programs aiming at a profession in the field of healthcare.

REFERENCES

- [1] Aalto, A., Korpilahti, U., Sainio, P., Malmivaara, A., Koskinen, S., Saarni, S., Valkeinen, H., & Luoma, M. (2013). Aikuisten geneeriset elämänlaatumittarit terveysterveys- ja hyvinvointitutkimuksessa sekä terveysterveys- ja kuntoutuspalvelujen vaikutusten arvioinnissa. Terveysten ja hyvinvoinnin laitos (THL). http://www.thl.fi/toimia/tietokanta/media/files/suositus/2013/10/30/Toimia_QOL_suositus_131031.pdf
- [2] Ashutosh, J., Andrews, D., Fiksdal, A., Kumbamu, A., McCormick, J., Misitano, A., Nelsen, L., Ryu, E., Sheth, A., Wu, S., & Pathak, J. (2014). Comparative Analysis of Online Health Queries Originating From Personal Computers and Smart Devices on a Consumer Health Information Portal. *Journal of Medical Internet Research*, 16(7). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4115262/>
- [3] Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach L, Kaufman J, & Neer SM (1997). The screen for child anxiety related emotional disorders (SCARED): scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(4), 545–553.
- [4] Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(10), 1230–1236.
- [5] Bradley, M., & Lang, P. (1999a). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, USA.
- [6] Britz, Denny (2015). Convolutional neural network for text classification in Tensorflow. <https://github.com/dennybritz/cnn-text-classification-tf>
- [7] Brown, P., Halász, S., Goodall, C., Cochrane, D., Milano, P., & Allegra, J. (2010). The ngram chief complaint classifier: A novel method of automatically creating chief complaint classifiers based on international classification of diseases groupings. *Journal of Biomedical Informatics*, 43(2), 268–272.
- [8] Bögels, S., & Van Melick, M. (2004). The relationship between child-report, parent self-report, and partner report of perceived parental rearing behaviors and anxiety in children and parents. *Personality and Individual Differences*, 37, 1583-1596.
- [9] Califf, R. (2014). The Patient-Centered Outcomes Research Network: A national infrastructure for comparative effectiveness research. *North Carolina Medical Journal*, 75(3).
- [10] Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.
- [11] Eilola, T., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, 42(1), 134–140. <http://link.springer.com/article/10.3758/BRM.42.1.134>

- [12] EuroQol Group (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16:199–208.
- [13] Finnish Dependency Parser (2018). Turku Dependency Treebank statistical dependency parser. The Finnish statistical dependency parser trained on the Turku Dependency Treebank. <http://bionlp.utu.fi/finnish-parser.html>;
<https://github.com/TurkuNLP/Finlish-dep-parser>
- [14] Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, 36(1), 23–50. <https://doi.org/10.1093/applin/amt020>
- [15] Foley, Y., & Mullis, F. (2008). Interpreting children's human figure drawings: basic guidelines for school counselors. *Georgia School Counselors Association (GSCA) Journal*, Volume 1, No. 1, 2008. <https://files.eric.ed.gov/fulltext/EJ864355.pdf>
- [16] Hale, W., Raaijmakers, Q., Muris, P. & Meeus, W. (2005). Psychometric properties of the screen for child anxiety related emotional disorders (SCARED) in the general adolescent population. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(3), 283-290.
- [17] Hays, R., Sherbourne, C., & Mazel, R. (1993). The RAND 36-item Health Survey 1.0. *Health Economics* 1993; 2:217–227.
- [18] Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, 20(10), 1727–1736.
- [19] Higginbotham, G., Munby, I., & Racine, J. (2015). A Japanese word association database of English. *Vocabulary Learning and Instruction*, 4(2), 1-20. doi: 10.7820/vli.v04.2.higginbotham.et.al
- [20] ICD-10 (2011). International Classification of Diseases (ICD-10), in Finnish. WHO and THL. <http://urn.fi/URN:NBN:fi-fe201205085423>
- [21] ICF (2013). International Classification of Functioning, Disability and Health (ICF), in Finnish. WHO and THL. <http://urn.fi/URN:NBN:fi-fe201303252595>
- [22] KAVI (2018). National Audiovisual Institute in Finland (KAVI), Department for Media Education and Audiovisual Media (MEKU). Classification of audiovisual programmes (criteria concerning age limits). <https://kavi.fi/en/media-education-and-audiovisual-media/classification-audiovisual-programmes>
- [23] Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. Association for Computational Linguistics. <http://www.aclweb.org/anthology/D14-1181>
- [24] Koskinen, S., Lundqvist, A., & Ristiluoma, N. (toim.) (2012). *Terveys, toimintakyky ja hyvinvointi Suomessa 2011 (Health, functional capacity and welfare in Finland in 2011)*. National Institute for Health and Welfare in Finland (THL), report 68/2012, Helsinki 2012. <http://urn.fi/URN:ISBN:978-952-245-769-1>
- [25] Lahti, Lauri (2016). Semantic modeling of healthcare guidelines to support health literacy and patient engagement. *Proc. Global Learn 2016: Global Conference on Learning and Technology*, 28-29 April 2016, Limerick, Ireland, pages 298-304, AACE. ISBN 978-1-939797-23-0. <https://www.learntechlib.org/p/172737/> (Open access: <http://urn.fi/URN:NBN:fi:aalto-201603291477>)
- [26] Lahti, Lauri (2017). Interpretation of health-related expressions and dialogues: enabling personalized care with contextual measuring and machine learning. *International Journal of New Technology and Research (IJNTR)*, volume 3, issue 11, November 2017, pages 171-179, ISSN 2454-4116. https://www.ijntr.org/download_data/IJNTR03110081.pdf (Open access: <http://urn.fi/URN:NBN:fi:aalto-201712298340>)
- [27] Lauri, Lahti (2018a). Expression collection of everyday life (In Finnish: Arkielämän ilmaisukokoelma), version as of December 2018. Department of Computer Science, Aalto University School of Science, Helsinki, Finland. To appear online at <http://aaltodoc.aalto.fi>
- [28] Lahti (2018b). Data supplement to the research article Lahti, Lauri, "Supporting care by interpretation of expressions about patient experience with machine learning", Aalto University School of Science, Finland. To appear online at <http://aaltodoc.aalto.fi>.
- [29] Lahti, Lauri, Tenhunen, Henni, & Nieminen, Marko (2018). How patients talk about care? Identifying patient experience expressions from online discussions. *Proc. Medical Informatics Europe (MIE 2018)*. ISBN 978-1-61499-851-8. <http://ebooks.iospress.nl/volumearticle/48765>
- [30] Lang, P., Bradley, M., & Cuthbert, B. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.
- [31] Marmot, M., Banks, J., Blundell, R., Lessof, C., & Nazroo, J. (eds.) (2003). *Health, wealth and lifestyles of the older population in England: The 2002 English longitudinal study of ageing*. ISBN: 1-903274-34-6; Institute for Fiscal Studies, December 2003. <http://www.elsa-project.ac.uk/reportWave1>
- [32] Nosikov, A., & Gudex, C. (eds.) (2003). EUROHIS: Developing common instruments for health surveys. *Biomedical and Health Research*, volume 57. IOS Press, Amsterdam, The Netherlands. http://www.euro.who.int/__data/assets/pdf_file/0015/101193/WA9502003EU.pdf
- [33] Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., & Terwee, C. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147-1157.
- [34] Samson, A., Kreibitz, S., Soderstrom, B., Wade, A., & Gross, J. (2016). Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and Emotion*, 30(5), 827-856. <https://doi.org/10.1080/02699931.2015.1031089>
- [35] Sintonen, H. (1994). The 15D measure of health related quality of life. II Feasibility, reliability and validity of its valuation system. Working paper 42. 1994. Centre for Health Program Evaluation, Australia.
- [36] Solans, M., Pane, S., Estrada, M., Serra-Sutton, V., Berra, S., Herdman, M., Alonso, J., & Rajmil, L. (2008). Health-related quality of life measurement in children and adolescents: a systemic review of generic and disease-specific instruments. *Value Health* 2008; 11(4), 742–764.
- [37] Spitzer, R., Kroenke, K., Williams, J., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097. <https://www.ncbi.nlm.nih.gov/pubmed/16717171>
- [38] Söderholm, C., Häyry, E., Laine, M., & Karrasch, M. (2013) Valence and arousal ratings for 420 Finnish nouns by age and gender. *PLoS ONE*, 8(8): e72859. <https://doi.org/10.1371/journal.pone.0072859>
- [39] de Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine*. Cambridge University Press, Cambridge, UK.
- [40] WHO (1997). World Health Organization (WHO). WHOQOL measuring quality of life. Division of Mental Health and Prevention of Substance Abuse. http://www.who.int/mental_health/media/68.pdf
- [41] Terveyskirjasto (2018). Healthcare guidelines given by Terveyskirjasto provided by The Finnish Medical Society Duodecim (in Finnish). <http://www.terveyskirjasto.fi>
- [42] The Suomi 24 Corpus (2016). The corpus of discussion forums of the Suomi24.fi website, provided by the Language Bank of Finland and CSC - IT Center for Science. <http://urn.fi/urn:nbn:fi:lb-2017021630>
- [43] Tsai, T., Lee, S., & Tsai, Y. (2015). Explaining selected health behaviors in a national sample of Taiwanese adults. *Health Promotion International*, 30(3), 563-572.

- [44] Warriner, A., Kuperman, V., Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191-207. <http://link.springer.com/article/10.3758/s13428-012-0314-x>