Semantic modeling of healthcare guidelines to support health literacy and patient engagement

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

Abstract: Developing new methods and solutions of personalized medicine can address many current socio-economical challenges both locally and globally. The investments made to support health can help people to have an independent, productive and happy life. To motivate the development of new patient support tools we illustrate the need for better health literacy and patient engagement, some common frameworks for modeling medical knowledge and some ways to support patients with online health queries and shared decision making. Then we provide some experimental results we have generated by semantic analysis about healthcare guidelines offered by The Finnish Medical Society Duodecim containing 85,055 words so that we created a conceptual network of 57,679 unique conceptual links traversed with 200,000 link steps. We suggest that our approach to semantic modeling of medical knowledge can be modularly applied to develop varied computational solutions for personalized medicine and health informatics.

1 Introduction

The World Health Organization's policy brief emphasizes the importance of supporting patient engagement to increase clinical and economic effectiveness and satisfaction as well as lists several distinctive roles of a patient that need to be addressed to enable a more patient-centered health care, including understanding the causes, self-diagnosing conditions, selecting and managing treatments, monitoring symptoms and effects, being aware of safety issues, learning to manage with a disease and adopting health behaviours (Coulter et al. 2008).

It has been estimated that about 80-90 percent of patients with long-term conditions and their carers can be supported to manage actively their own health (Da Silva 2011). It has been suggested that the expected growth of elderly patients should motivate moving focus from episodic treatments towards management of long-term conditions and creation of care plans that can increase a patients self-management and decrease emergency hospital admission but still it appears that well-structured care planning has been not been yet actively implemented (NHS England 2013).

2 Previous research

2.1 Need for better health literacy and patient engagement

A popular approach to explain self-regulation of motivation is based on Bandura's theory of self-efficacy that suggest that people form beliefs about their capacity that affects the feeling of control and confidence to manage challenges (Bandura 1984). In educational field it has been suggested that the development of self-efficacy can be supported by for example setting performance and learning goals, getting progress feedback and reaching a positive loop of empowerment (Scott 1996).

Based on an analysis of health survey data it has been identified that moderate correlations exist among health literacy, self-efficacy (perceived ability to improve and maintain personal health) and health locus of control (perceived control over personal health behaviour and status) so that they are independent but correlated factors (Tsai et al. 2015). People having a low health literacy have been identified to have poorer health status, higher rate of hospital admissions, more rejecting treatments and care plans, get more drug and treatment errors and use less preventive services (Board on Neuroscience and Behavioral Health 2004).

It has been considered that autonomous voluntary self-help groups do not have capability to replace professional health services but can complement it and mobilize new resources to provide health care, and thus there is a need to get sufficient support for these groups from the medical professionals and organizations (Nayar et al. 2004). It has been estimated that self-help groups reach about 6-9 percent of potential participants, so that males, minorities, the aged and the working and lower classes are under-represented (Nayar et al. 2004). In an evaluation of Internet-based self-care technology it was found that patients experienced challenges about navigation and search functionalities whereas caregivers about lack of feedback and documentation possibilities, further challenges included insufficient tailoring of information and personalization of advice as well as language, reliability and legal
or ethical issues (Nijland et al. 2008).

It is ethically important that in health care that all persons affected by medical diagnosis and decision making need to have a possibility to be well heard thus respecting the patient's rights and also it is important to recognize that public perspectives towards medical work are complex, dynamic and relational (Schicktanz et al. 2012). With an effort to prevent harms caused to patients in health care by medical errors one initiative has been to accumulate medical error data to form an ontology and based on eighth taxonomies it was possible to identify 12 multidimensional axes representing perspectives of medical error events, including practitioners involved, patient profile, health care service, error location, contributing factors, professional activity, time of error, system factors, patient outcome, human factors, interface design factors, medical product involved (Mokkarala et al. 2008).

The quality of communication between a patient and a doctor has been shown to relate with better identification of the patient's needs (Ha & Longnecker 2010) and the sense of control has been associated for example with the ability to manage with pain, to recover from illness, to have decreased tumor growth and to manage with daily life (Roter 1983; Greenfield et al. 1985; Greenfield et al. 1988). It has been argued that perspectives of patients should be better highlighted and integrated in health communication and solutions have been suggested based on perspectivist and polyocular theories that aim to assist communication and learning by making concluding analysis about meanings, values and interests of complementing perspective groups (Fage-Butler 2013). It has been suggested that to achieve a multidimensional space of understanding the observations made by first-order scientific perspectives about a research object need to be observed by second-order observations that have sufficiently separate resources (Alrøe & Noe 2011).

2.2 Frameworks for modeling medical knowledge

It has been suggested that making medical research more open can enable better transparency and validation of results and methods that also encourages publishing both positive and negative results as well as can enable having open data that is normed, machine readable and in standardized format (Floca 2014). There have emerged open collaborative initiatives to develop medical diagnostic tools and methods that can be freely shared and used (for example OpenClinical community aims to promote use of knowledge management technologies for clinical research and patient care (http://www.openclinical.org) and OpenMRS community aims to develop enterprise-level open source medical records systems (http://www.openmrs.org)).

It has been suggested that the design and function of medical decision support systems can be categorized based on 24 features that belong to five groups which are context, knowledge and data source, decision support, information delivery and workflow (Sim & Berlin 2012). To improve accuracy of medical diagnostics that traditionally relies on classifying measurements along exemplary values it has been suggested that measurements can be mapped to vectors in conceptual spaces which have specifically defined relationships to an ontology and this approach has enabled successfully connecting medical sensor measurements and symbolic medical data (Dietze et al. 2009). Several ontology-related formalizations suitable for semantic modeling of medical texts have been developed and applied in the field of medical terminology and health care, including for example SNOMED CT (i.e. Systematized Nomenclature of Medicine - Clinical Terms) (IHTSDO 2016), ICD-10 (i.e. International Statistical Classification of Diseases and Related Health Problems, 10th revision) (WHO 2016), MeSH (i.e. Medical Subject Headings) (U.S. National Library of Medicine 2016), Health Level-7 (HL7) Clinical Statement Model, Standards-Based Active Guideline Environment (SAGE) and Protégé ontology editor.

Frameworks have been proposed to enable easier development of medical diagnosis decision support systems based on domain requirements modeling (Amarakoon et al. 2012). Previous research efforts have accumulated collections of medical algorithms that can be used for example as support tools to make diagnoses. For example The Medical Algorithms Company provides online access to over 20000 medical tools (http://www.medicalalgorithms.com/). Anyway many potential sources for challenges and errors can be encountered when trying to apply medical algorithms for actual diagnosis and treatment work (Johnson et al. 2002). There also exists online services that offer probabilistic suggestions for making a medical diagnosis based on a patient dataset (for example Promedas online service at http://www.promedas.nl). Some promising research initiatives rely on collaborative open source software projects, such as The Open Cognition Project that aims to create artificial intelligence that could reach the cognitive capacities of a human (http://wiki.opencog.org).

In Finnish context respected health information resources are provided by The Finnish Medical Society Duodecim. Duodecim's popular health information services for ordinary people is Terveyskirjasto (http://www.terveyskirjasto.fi) and for medical professionals Terveysportti (http://www.terveysportti.fi) and Oppiportti (http://www.oppiportti.fi). Duodecim is developing a clinical decision support system EBMeds (The Evidence-Based Medicine electronic Decision Support; http://www.ebmeds.org) that is in a collaboration with non-
2.3 Supporting patients with online health queries and shared decision making

A study indicated that the motivation for making online searches for health information can be categorized into symptom troubleshooting, searching to enhance a clinic visit and making searches for someone else (Fiksdal et al. 2014). In respect to the online health information searching strategy, people typically carried out a process of comparing and contrasting various sources against a personalized criteria relying on needs and reputation so that this filtering can finish when the results become repetitive or an exhaustion is experienced (Fiksdal et al. 2014). In a study examining the elements of clinical interactions in psychiatric care enabling shared decision making it turned out that a significal predictor for agreement in clinical decisions was incorporating consumer preference in discussions so that the probability of a full agreement was four times higher when the exploration of consumer preference was discussed more completely (Fukui et al. 2014).

Patient decision aids have been developed as methods to help individuals to manage the process of choosing between healthcare options, especially in preference-sensitive areas of healthcare and some characteristics used to evaluate them include universality (accessibility for diverse groups of people) and usability (ease of learning, efficiency of use, memorability, error frequency and satisfaction) (Hoffman et al. 2013). A review of studies evaluating delivering patient decision aids on the Internet indicated that using these methods improved knowledge measures, provided similar ratings for effort, convenience and satisfaction, and had variable effects on preferences about screening (Hoffman et al. 2013). Another review of studies indicated that using these methods improved knowledge, preparation for decision making and decisional conflict measures, was acceptable to patients, and had variable effects on preference about treatment or probability of receiving screening (Hoffman et al. 2013).

It has been suggested that future research about patient decision aids should analyze properties of interactivity, personally-tailored information, interpersonal communication, patient stories and socially-generated experiential information as well as evaluate how various features can improve the patient's sense of preparation, self-efficacy, decision quality and retaining and transferring decision-making skills (Hoffman et al. 2013). Furthermore in the development of patient decision aids evaluation should be made concerning features helping comprehension, risk-benefit comparison, strategies to assess if patient is well-informed, patient activation, diverse communication, coupling of information sources and clinical contextualization (for example decisions in preventive, acute, chronic and end-of-life treatment) as well as supporting various actors in medical context, accessibility, multiple evolving devices, social connectivity and privacy (Hoffman et al. 2013).

It has been suggested that medical community should pay more attention to information gathered independently by patients and its motivation (Dutta-Bergman 2005). It has been suggested that enabling access to health information through the Internet can transform the authoritative role of medical professionals to be more collaborative with the patient in the form of shared decision making (Winker et al. 2000). When translating evidence-based, population-based estimates about risks and benefits to the health care case of an individual patient there are many uncertainties (Politi et al. 2013). It has been found that patients are more satisfied with their care and relationship with caretakers if data and its uncertainties are expressed and managed openly (Parascandola et al. 2002).

It has been suggested that applying shared decision making approach can help to decrease unintentional variations in preference-sensitive clinical practice and thus to improve health on both individual and public level (Politi et al. 2013). It has been suggested that decision support intervention can supplement conversations in preference-sensitive care concerning practice guidelines for health promotion and disease prevention if the intervention enables linking a personalized decision-making approach with the clinical encounter (Politi et al. 2013). For example decision support interventions explaining the limitations of evidence in respect to certain cases can help the patient to achieve deliberation about options and then during a clinical visit it is possible to offer choices, summarize options, check understanding, elicit preferences and provide time for reviewing options (Politi et al. 2013).
3 Method

In our previous work (Lahti 2015c) we generated a conceptual co-occurrence network for a set of 3018 highest-ranking nouns of British National Corpus (BNC) and thus we gained altogether a set of 54 610 unique pairs of nouns, containing 2994 of 3018 unique nouns. Our conceptual co-occurrence network was generated based on the online database of British National Corpus (Leech et al. 2001) and the online database for Google Web 1T 5-gram database (FAU Erlangen-Nürnberg 2015).

In the latest work (Lahti 2016a) we focused on a subset of the conceptual co-occurrence network between 3018 highest-ranking nouns (as discussed in Lahti (2015c)) formulated by matching with Medical Academic Word List (MAWL) that is a medical vocabulary defined based on the most frequently used medical academic words in a set of medical research articles having 1 093 011 running words (Wang et al. 2008). It turned out that 257 of 623 unique concepts of the Medical Academic Word List occurred in our conceptual co-occurrence network. In the subset of the original conceptual co-occurrence network we observed those 1005 unique potentially explorable concept pairs that had nouns of Medical Academic Word List (MAWL) and had the highest rankings in British National Corpus (BNC) when considering the sum of ranking values of both concepts of the concept pair. We gained altogether 9501 shortest paths connecting the set of 1005 unique potentially explorable concept pairs thus having for each concept pair on average 9.45 alternative shortest paths of the shortest length and the length of the shortest paths was on average 2.98 link steps and the median value was 3 link steps.

Motivated by the previous research we have now made an analysis based on a set of 93 medical texts about healthcare guidelines given by Terveyskirjasto provided by The Finnish Medical Society Duodecim ("Käypä hoito, potilasversiot"; retrieved in January 2016 from http://www.terveyskirjasto.fi/terveyskirjasto/tk.koti?p_osio=109&p_teos=khp) including altogether 85 055 words. The analysis was made in Finnish but we now report the results in English. We found out that this set of texts contained 2014 unique nouns having at least 3 occurrences in the texts and these 2014 unique nouns had altogether 101 024 co-occurrences in shared sentences. Among these 101 024 co-occurrences there were 28 840 unique concept pairs. When each of 28 840 unique concept pairs was considered to form two conceptual links going into the opposite directions between these two concepts, we gained a relatively comprehensive conceptual co-occurrence network about health-related topics containing 57 679 unique conceptual links.

Table 1 shows a listing of the nouns in 93 medical texts about healthcare guidelines given by Terveyskirjasto along the flow of texts. Table 2 shows the highest-ranking nouns in 93 medical texts about healthcare guidelines given by Terveyskirjasto. Table 3 shows the highest-ranking traversed links along the random exploration path of 200 000 traversed link steps in the conceptual co-occurrence network based on 93 medical texts about healthcare guidelines given by Terveyskirjasto. Table 3 shows also the highest-ranking traversed concepts along this same random exploration path of 200 000 traversed link steps. We suggest that our method and results show such an approach to semantic modeling of medical knowledge that can be modularly applied for developing varied computational solutions for personalized medicine and health informatics. We want to enhance our contribution to the development of new patient support tools by publishing as open data an extensive listing of our semantic analysis results illustrated here only partially (the data is downloadable from Lahti (2016c)).

4 Discussion and future work

For a long time medical organizations have collected large databases about personal and collective medical histories and there have been attempts to analyze these knowledge entities with innovative methods to reveal unknown dependencies and causalities between living habits, symptoms, diseases, treatment and healing. Along the years it has become increasingly cost-efficient to store, maintain and access large data collections both at organizational and personal level. However, the amount of stored information has been constantly growing and there is a need to develop solutions to manage large data collections efficiently. These new solutions are needed to help to identify for example hidden features of dependency and causality as well as typical repeating patterns and trends.

The recent progress of digitalization and online networking of the society has opened new possibilities also for ordinary people for collecting, maintaining, synthesizing, accessing and sharing information and knowledge. These possibilities have also changed how ordinary people can manage their personal everyday life and collectively contribute to development of the society in various domains of life. Also in respect to better understand and manage personal health there have appeared new possibilities to gather and analyze information due to the increased easy online access to large up-to-date information resources and due to the introduction of personal intelligent mobile healthcare devices and software solutions. There is a strong need to develop intuitive computational tools to manage and analyze health data.
Table 3. a) Highest-ranking traversed links along the random exploration path of 200,000 traversed link steps in the conceptual co-occurrence network of 57,679 unique links containing 2014 highest-ranking nouns in 93 medical texts about healthcare guidelines given by Terveyskirjasto. A traversed link proceeds from ConceptA to ConceptB. b) Highest-ranking traversed concepts along this same random exploration path of 200,000 traversed link steps.

<table>
<thead>
<tr>
<th>Noun</th>
<th>Freq</th>
<th>ConceptA</th>
<th>ConceptB</th>
<th>Freq</th>
<th>ConceptA</th>
<th>ConceptB</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>1328</td>
<td>medication</td>
<td>168</td>
<td>amount</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>patient</td>
<td>685</td>
<td>need</td>
<td>165</td>
<td>adolescent</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>symptom</td>
<td>567</td>
<td>part</td>
<td>164</td>
<td>heart</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>child</td>
<td>389</td>
<td>follow-up</td>
<td>163</td>
<td>form</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>patientversion</td>
<td>318</td>
<td>physical exercise</td>
<td>159</td>
<td>epilepsy</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>275</td>
<td>percent</td>
<td>156</td>
<td>change</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medicine</td>
<td>262</td>
<td>disease</td>
<td>155</td>
<td>text</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>year</td>
<td>244</td>
<td>woman</td>
<td>154</td>
<td>prevention</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trace</td>
<td>237</td>
<td>diabetes</td>
<td>151</td>
<td>reporter</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disease</td>
<td>232</td>
<td>week</td>
<td>144</td>
<td>sigh</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>example</td>
<td>205</td>
<td>finish</td>
<td>143</td>
<td>wound</td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pain</td>
<td>201</td>
<td>smoking</td>
<td>132</td>
<td>asthma</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reason</td>
<td>198</td>
<td>month</td>
<td>130</td>
<td>skin</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>month</td>
<td>198</td>
<td>blood pressure</td>
<td>130</td>
<td>problem</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>surgery</td>
<td>198</td>
<td>help</td>
<td>129</td>
<td>seizure</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>research</td>
<td>192</td>
<td>rehabilitation</td>
<td>128</td>
<td>diabetic</td>
<td>87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medical society</td>
<td>191</td>
<td>risk factor</td>
<td>124</td>
<td>motive</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>supplement</td>
<td>190</td>
<td>inflammation</td>
<td>122</td>
<td>half</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medication</td>
<td>181</td>
<td>pregnancy</td>
<td>121</td>
<td>day</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>teeth</td>
<td>180</td>
<td>month</td>
<td>119</td>
<td>situation</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>use</td>
<td>179</td>
<td>specialized doctor</td>
<td>110</td>
<td>activity</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>doctor</td>
<td>171</td>
<td>age</td>
<td>107</td>
<td>infection</td>
<td>85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk</td>
<td>169</td>
<td>connection</td>
<td>106</td>
<td>person</td>
<td>84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


Board on Neuroscience and Behavioral Health (2004). Health literacy: a prescription to end confusion. Board on Neuroscience
and Behavioral Health, Institute of Medicine. National Academies Press, Washington, DC, USA.
Coulter, A., Parsons, S., & Askham, J. (eds.) (2008). Where are the patients in decision-making about their own care? World
Health Organization, Policy brief: Health Systems and Policy Analysis. WHO Regional Office for Europe and European
Observatory on Health Systems and Policies.
Da Silva, D. (2011). Helping People Help Themselves: A review of the evidence considering whether it is worthwhile to support
Dutta-Bergman, M. (2005). Developing a profile of consumer intention to seek out additional information beyond a doctor: the
role of communicative and motivation variables. Health Communication, 17(1), 1-16.
http://corpora.linguistik.uni-erlangen.de/demos/cgi-bin/Web1T5/Web1T5_colloc perl
Online Health Information Searching: A Qualitative Approach to Exploring Consumer Perspectives. Journal of Medical Internet
Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing. ISBN: 978-3-319-00025-1
(Print) 978-3-319-00026-8 (Online).
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3984060/
Internal Medicine, 102(4), 520–528.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3096184
patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. BMC Medical
IHTSDO (2016). International Health Terminology Standards Development Organisation. Systematized Nomenclature of
Medicine – Clinical Terms (SNOMED CT). http://www.ihtsdo.org/snomed-ct/
Johnson, Kathy A.; Swirbely, John R.; Sriram, M.G.; Smith, Jack W.; Kantor, Gareth; Rodriguez, Jorge Raul (November 2002).
"Automated Medical Algorithms: Issues for Medical Errors". Journal of the American Medical Informatics Association 9 (6 Suppl
Lahti, Lauri (2015c). Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language
ability levels and frequency ranking. Proc. E-Learn - World Conference on E-Learning, 19-22 October 2015, Kona, Hawaii, USA.
Online: http://urn.fi/URN:NBN:fi:aalto-201509294488
5th International Conference on Human Computing, Education & Information Management System (ICHCEIMS 2016), 27-28
March 2016, Sydney, Australia. (Open access: http://aaltodoc.aalto.fi/)
Lahti, Lauri (2016c). Supplement to Lauri Lahti’s conference article “Semantic modeling of healthcare guidelines to support
ability levels and frequency ranking. Proc. E-Learn - World Conference on E-Learning, 19-22 October 2015, Kona, Hawaii, USA.
Johnson, Kathy A.; Swirbely, John R.; Sriram, M.G.; Smith, Jack W.; Kantor, Gareth; Rodriguez, Jorge Raul (November 2002).
"Automated Medical Algorithms: Issues for Medical Errors". Journal of the American Medical Informatics Association 9 (6 Suppl
Lahti, Lauri (2015c). Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language
ability levels and frequency ranking. Proc. E-Learn - World Conference on E-Learning, 19-22 October 2015, Kona, Hawaii, USA.
Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, USA. http://www.editlib.org/p/151985/
"Automated Medical Algorithms: Issues for Medical Errors". Journal of the American Medical Informatics Association 9 (6 Suppl
5th International Conference on Human Computing, Education & Information Management System (ICHCEIMS 2016), 27-28
March 2016, Sydney, Australia. (Open access in Aaltodoc publication archive at http://aaltodoc.aalto.fi/)
Lauri, Lahti, Lauri (2016c). Supplement to Lauri Lahti's conference article "Semantic modeling of healthcare guidelines to support
http://ucrel.lancs.ac.uk/bncfreq/lists/5_1_all_rank_noun.txt)
Ontology. In Henriksen, K., Battles, J., Keyes, M. et al. (eds.), Advances in Patient Safety: New Directions and Alternative
NHS England (2013). High quality care for all, now and for future generations: Transforming urgent and emergency care services
Supporting Self-Care: Problems Encountered by Patients and Caregivers When Using Self-Care Applications. Journal of Medical Internet Research, 10(2).


