Interactive health insight miner: an adaptive, semantic-based approach

Isabel Funke
isabel.funke@gmail.com

Rim Helaoui and **Aki Härmä**
Philips Research
High Tech Campus 34
5656 AE Eindhoven
The Netherlands
rim.helaoui@philips.com
aki.harma@philips.com

Abstract

E-health applications aim to support the user in adopting healthy habits. An important feature is to provide insights into the user's lifestyle. To actively engage the user in the insight mining process, we propose an ontology-based framework with a Controlled Natural Language interface, which enables the user to ask for specific insights and to customize personal information.

1 Introduction

E-health services based on wearable sensors, such as smart watches, need methods to discover insights from the sensor data. *Insights* describe user-specific behavior patterns, or habits, that are relevant for guiding the user towards a healthy lifestyle. For example, an insight might reveal that the user is especially sedentary at the weekend.

Blind discovery of significant insights is essentially a search problem and requires a lot of data. If the discovery of insights took place in dialogue with the user, the search problem could be restricted to areas that interest the user the most. Also, the user could provide complementary information that cannot be inferred from the data.

In this paper, we propose a description logics-based approach towards an interactive system for the discovery of insights. Concretely, we describe an ontological framework implemented on top of a statistical insight miner (Härmä and Helaoui, 2016) that enables the natural language-based retrieval and customization of insights from wearable sensor data.

2 Proposed framework

Our framework consists of five layers, see Fig. 1. Data is acquired in the *data layer* and further processed in the *information extraction layer*. The *information & knowledge integration layer* abstracts the extracted information into formal *facts*. The resulting knowledge base can include user- and situation-specific information as well as common sense knowledge. The *reasoning layer* leverages logic-based algorithms that reason with the available knowledge. The *verbalization layer* transforms the facts into coherent and comprehensible natural language (NL) messages. Similar systems for *data-to-text* summarization have been proposed in the literature (e.g. Portet et al., 2009).

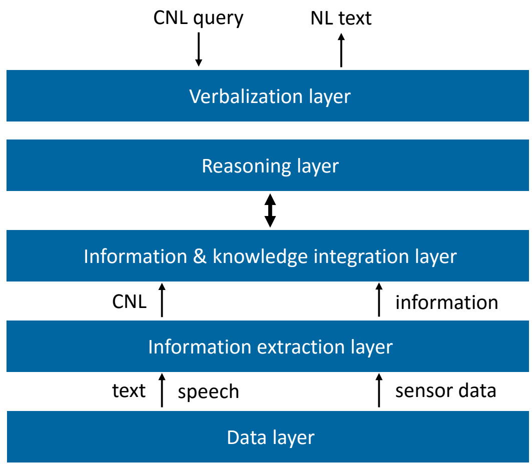

Figure 1: Proposed framework for the interactive health insight miner.

We additionally introduce a *Controlled Natural Language (CNL)*. It is a formal language that can be translated unambiguously into knowledge base facts, but is also understandable by humans. By adopting the CNL, the user can interact with the system, i.e., add and query facts from the knowledge base. Natural language or spoken text can be fed into the system after translation into CNL.

3 Representing, summarizing and verbalizing insights

The user's lifestyle is described by an *ontology* that contains the routines, habits, and targets of the

Proceedings of The 11th International Natural Language Generation Conference, pages 478–479,
Tilburg, The Netherlands, November 5-8, 2018. ©2018 Association for Computational Linguistics

user. These concepts are leveraged to represent insights as knowledge base facts.

Inspired by NaturalOWL (Galanis and Androutsopoulos, 2007), we include *lexical annotations* in the ontology, which specify how ontology concepts are to be translated into natural text. This way, the ontology also acts as a lexicon. We include the lexical categories (e.g., noun, determiner, preposition, or verb) in the annotations to facilitate the use of standard *Natural Language Generation (NLG)* techniques, such as adapting verb conjugations, adapting the verb tense, or aggregating sentence parts.

To enable user interaction, we specify a CNL based on the vocabulary defined in the ontology. The CNL plays the role of a human and machine understandable interface which allows to directly map the user's input to the formal concepts of the ontology. This way, the user can add personal information to the system, e.g., "On Monday at work, I play tennis". This statement will be formalized as a fact and added to the knowledge base. The CNL also provides the basis for verbalizing the system's responses to the user's queries, such as "What are insights about Sunday afternoon?" We use the *Backus-Naur* form to specify the CNL as a context-free grammar.

To create an NL summary of a number of insights, we implement the following NLG steps (Bouayad-Agha et al., 2014): (1) *Content selection*: We let the user ask for specific insights, for example insights about their step count on Sunday. (2) *Discourse planning*: We group those insights together that are semantically related. The insights are first grouped by the measurement to which they refer (e.g. step count) and then ordered within each group from more general to more specific. For the grouping and ordering steps, we leverage our semantic model (ontology) and apply reasoning algorithms to determine which relationships hold between which insights. (3) *Lexicalization*: We follow a template-based approach using the lexical annotations in the ontology. (4) *Aggregation*: We verbalize each group of equally specific insights using an aggregation template. Finally, for (5) *realization*, we use the realization engine SimpleNLG (Gatt and Reiter, 2009).

4 Implementation

We demonstrate the viability of our framework by implementing one use case related to lifestyle insight mining. The ontology is implemented in OWL using the Protégé[1] editor. We use OWL API[2] 4.1.3 together with the reasoner HermiT[3] 1.3.8 to implement our system in Java. To implement the verbalization functionality, we build upon the library SimpleNLG[4].

5 Example run

> What are insights about my sedentary time on Sunday?

> *On Sunday your sedentary time is lower than on Tuesday, Friday, Wednesday or Thursday.*
On Sunday in the afternoon your sedentary time is lower than on Friday, Wednesday, Thursday or Tuesday in the afternoon.

> What are insights about when I go running?

> *Could not find any insights about this.*

> On Sunday in the afternoon, I go running.

> What are insights about when I go running?

> *On Sunday in the afternoon when you go running your sedentary time is lower than on Tuesday, Friday, Thursday or Wednesday in the afternoon.*

References

Nadjet Bouayad-Agha, Gerard Casamayor, and Leo Wanner. 2014. Natural language generation in the context of the semantic web. *Semantic Web*, 5(6):493–513.

Dimitrios Galanis and Ion Androutsopoulos. 2007. Generating multilingual descriptions from linguistically annotated OWL ontologies: the NaturalOWL system. In *Proceedings of the 11th European Workshop on Natural Language Generation*, pages 143–146.

Albert Gatt and Ehud Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93.

Aki Härmä and Rim Helaoui. 2016. Probabilistic scoring of validated insights for personal health services. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6.

François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816.

[1] http://protege.stanford.edu/
[2] https://github.com/owlcs/owlapi
[3] http://www.hermit-reasoner.com/
[4] https://github.com/simplenlg/simplenlg