Personal Health and Fitness Analysis Using Semantic Web and Embedded Systems

Mark A. Swanson

Florida Atlantic University

Abstract

Fitness bands and other personal health tracking devices have become commonplace for people of all health and lifestyle backgrounds. The completed project described herein collects tracking data from an existing fitness band product and combines this data with ambient environment data gathered from an embedded system which would reside in a consumer's residence. Storage and analysis of the data occurs locally on the embedded system. Simple analysis is performed which relates the two data sets utilizing the Semantic Web stack; an example analysis categorizes a given day as either 'Lazy' or 'Active'. The combination of a hardware ecosystem, physical monitoring of both the consumer and the consumer's environment, and finally the use of Semantic Web compatible solutions for data analysis present a complete Internet of Things solution for personal health and fitness analysis and provides a platform for easy future integration into the Semantic Web.

Keywords: Semantic Web, Raspberry Pi, Health, Fitness

Personal Health and Fitness Analysis Using Semantic Web and Embedded Systems

According to analyst firm Canalys, seven million fitness bands were sold in the first quarter of 2015 (“Media alert,” 2015). The data associated with these devices may already be stored in a cloud service either through a third-party application or by the company who sells the device. Smart watches, cell phones, and in-home products have begun tracking and storing personalized data which may be used for assisting in the prognosis or diagnosis of specific health metrics or diseases. Presently the majority of consumer-grade products are limited to displaying and storage of data with little analysis being done. Integrating various data sources and performing further analysis on them may lead to further insights into the well-being of an individual or communities.

Project Development

Project development was separated into distinct steps to enforce separation of project components and minimize reliance on any given step being completed prior to beginning or continuing another.

Raspberry Pi

A Raspberry Pi was utilized as the embedded system for both monitoring ambient conditions and combining then analyzing ambient data with fitness tracking data. The Raspberry Pi was selected as an embedded system platform due to its large and ever-growing community as well as its stable and reliable hardware ecosystem.

Ambient data was collected on the Raspberry Pi embedded system, refer to Figure 1, utilizing three jumper cables and the GPIO header available on the hardware platform. Python and the RPi.GPIO module were utilized to collect and store timestamp data for ambient environment events in a Comma Separated Values (CSV) file on the local memory of the Raspberry Pi.

Garmin© vívofit®

Collection of Garmin© vívofit® data was manually performed by downloading the data from Garmin's online service *Garmin Connect* which exported a CSV file for later integration with ambient data generated by Raspberry Pi. Loading the data on to the Raspberry Pi was done over a Wi-Fi connection while using the Raspberry Pi monitor. Daily step counts by a weekly basis were exported for use with the project. The week of July 19, 2015 was selected for sample analysis.

Data Integration

Data integration was performed on the Raspberry Pi system utilizing Python with csv and datetime modules. Separate CSV files for both ambient and fitness data were combined into one dataset, in CSV a file, for simpler human processing and potential future uses. A 3-tuple was utilized to store date, event, and values for each entry in the integrated CSV file for both timestamp event and daily step counts. Again, data was stored on the memory storage of the Raspberry Pi.

protégé

The RDF file for the combined ambient and fitness tracking data was created using protégé. The RDF file utilized the Web Ontology Language (OWL) (http://www.w3.org/TR/owl2-overview/), Resource Description Framework (RDF) (http://www.w3.org/TR/rdf11-concepts/), RDF Schema (RDFS) (http://www.w3.org/TR/rdf-schema/), and XML Schema Definition Language

(XSD) (http://www.w3.org/TR/xmlschema11-1/) as the base schemas and ontologies. Time Ontology in OWL (http://www.w3.org/TR/owl-time/) is an ontology of temporal concepts and was used for calculated duration data. Data was loaded manually for overall construction of the ontology. The built-in HermiT 1.3.8.3 reasoner was used for performing reasoning on the RDF file, refer to Figure 2.

Results

Three hardware interfaces were used for the collection of ambient data. Manual download from the manufacturer of the fitness band was used for the collection of fitness data. The Raspberry Pi generated the ambient data file, and processed both the ambient data and fitness data files to generate a third, integrated file for analysis. protégé, with aforementioned schemas and ontologies, was used to categorize the integrated data file into two separate categories for a given consumer's day (i.e. 'Lazy', 'Active'). For the given sample analysis week, using data from the author, three days were categorized as 'Active', and three days were categorized as 'Lazy'.

Discussion

Deeper analysis can be performed when combining multiple sources of health and fitness data. Adding qualifiers beyond what a fitness band can collect to the classification of a person's day garners further efficiency and accuracy than what is otherwise available standalone. Discovering classifications, trends, and formulating health hypotheses on individuals and communities may be enabled by a project similar to this one, but the analysis performed by this project was quite minimal. Further research should be performed in this area.

Ambient data is, presently, far more difficult to obtain as it often needs to be collected external to devices a consumer may wear or always have with them at all times. Other sources of ambient data may not be capable of communicating with external devices, such as the TV discussed herein. This interface from the ‘real’ (analog) world to the ‘web’ (digital) world is certainly an ever-evolving area, as seen by platforms such as the Raspberry Pi, but broadly available off-the-shelf goods which can assist with this form of data collection are still not prevalent enough today to harness the average consumer for large amounts of ambient data.

Conclusion

The creation of a Semantic Web capable health and fitness data integration platform was completed successfully. Through a combination of embedded systems, interpreted languages, reasoners, and occasionally manual input, a completed platform for future use can be recreated from this project. Health and fitness analysis was quite minimally performed, with ample opportunity for further study and research being available in this arena.

References

Chris, J., Nicole, P., Tim, C., & Rachel, L. (2015, May 20). *Media alert: Fitbit maintains leadership share of wearable band market before Apple Watch entrance.* Retrieved from http://www.canalys.com/newsroom/media-alert-fitbit-maintains-leadership-share-wearable-band-market-apple-watch-entrance

Appendix A

Online Resources

- Raspberry Pi: http://www.canakit.com/raspberry-pi/raspberry-pi-kits
- protégé: http://protege.stanford.edu/
- Programming, data, and screen shots: https://github.com/Semantic-Web/Mark-S/tree/master/Final

Figures

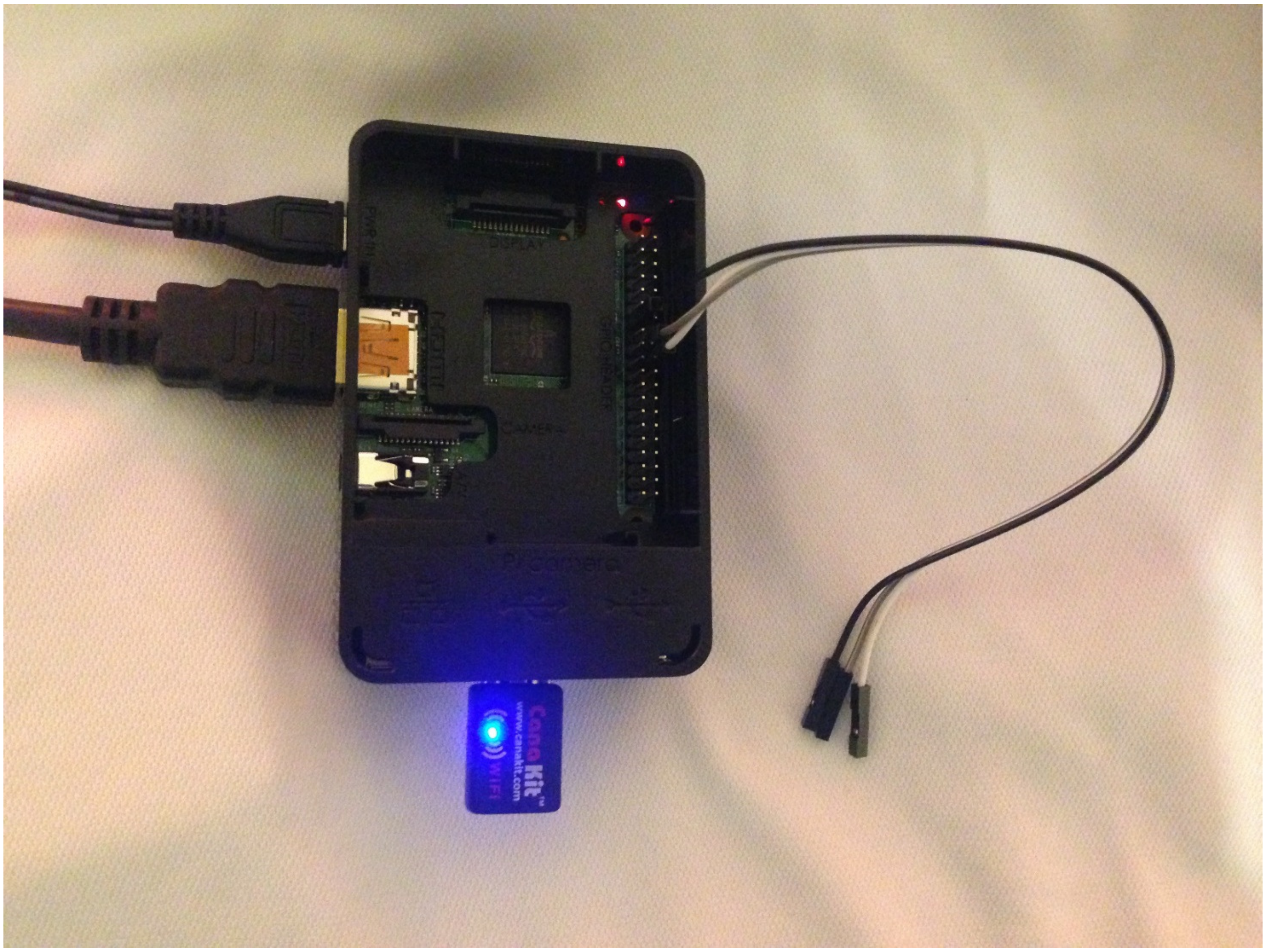

Figure 1. Raspberry Pi Embedded System Hardware Configuration for Timestamp Generation and Local File Storage and Processing

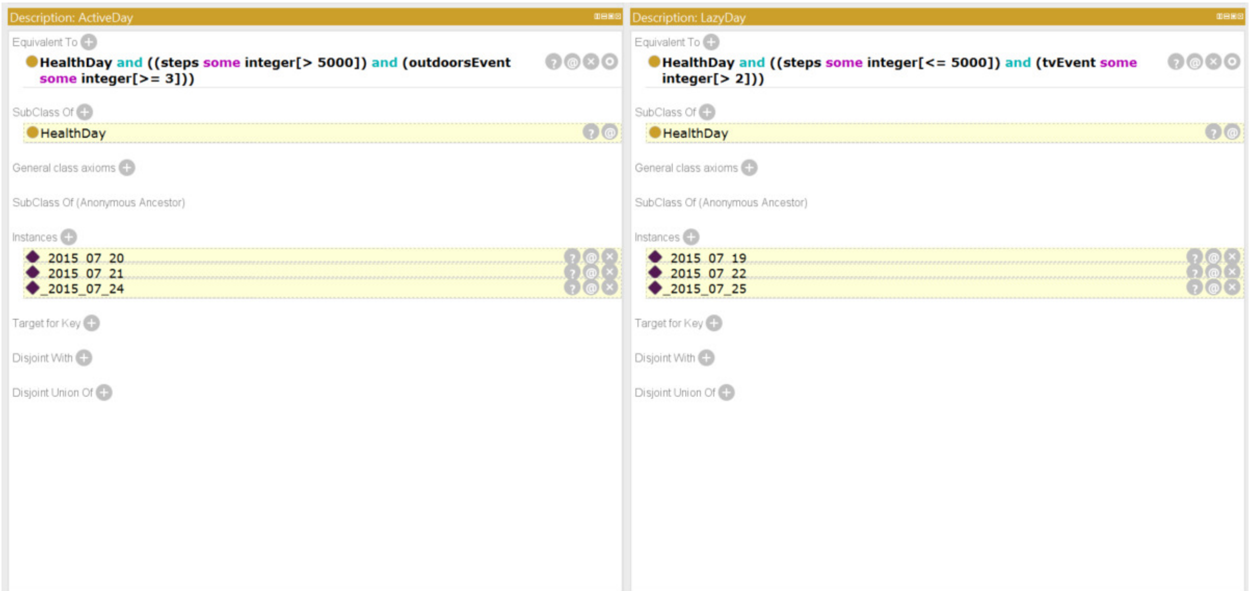

Figure 2. Results of RDF Analysis Utilizing protégé for ActiveDay and LazyDay Categorization Using Author Sample Data