Cardiovascular Health and Physical Activity:
A Model for Health Promotion and Decision Support Ontologies

Vimala Ponna
Department of Neurobiology, Physiology, and Behavior
University of California at Davis
Davis, CA
vmponna@ucdavis.edu

Aaron Baer
Department of Food Science and Technology
University of California at Davis
Davis, CA
ambaer@ucdavis.edu

Matthew Lange, PhD
Department of Food Science and Technology
University of California at Davis
Davis, CA
mclange@ucdavis.edu

***Abstract*— Current cardiovascular disease decision support systems (DSS) rely primarily on ontologies that characterize and quantify disease, recommending appropriate pharmacotherapy (PT) and/or surgical interventions (SI). PubMed and Google Scholar searches reveal no specific ontologies or literature related to DSS for recommending physical activity (PA) and diet interventions (DI) for cardiovascular health and fitness (CVHF) improvement. This dearth of CVHF-PA/DI structured knowledge repositories has resulted in a scarcity of user-friendly tools for scientifically validated information retrieval about CVHF improvement. Advancement of health science depends on timely development and implementation of health (rather than disease) ontologies. We developed a time-efficient workflow for constructing/maintaining structured knowledge repositories capable of providing informational underpinnings for CVHF-PA/DI ontologies and DSS that support health promotion, including precise, personalized exercise prescription. This workflow creates conceptual lattices about effects of varied PA on CVHF. These conceptual maps lay the foundation for accelerated creation of health-focused ontologies, which ultimately equip DSS with CVHF knowledge related PA and DI.**

INTRODUCTION

Current healthcare ontologies and DSS rely primarily on knowledge relevant to disease risk assessment and treatment and are focused almost entirely on assessing PT and SI. Analogous ontologies and DSS for advancing consumer health via PA do not yet exist. Successful implementation of healthcare ontologies and DSS for recommending specific PT and SI for cardiovascular diseases are built upon databases from clinical trials and patient records, combined with highly curated, hierarchical vocabularies of diseases, diagnoses, PT, and SI. PubMed and Google Scholar searches reveal no scientific literature about healthcare ontologies and consumer DSS for CVHF using analogous systems related to knowledge about PA and DI for health improvement. Medicine today relies heavily on modeling disease, rather than modeling health. Part of the problem is the dearth of queryable, curated and structured knowledge repositories dedicated to CVHF relative to specific DI and PAs. These immense reserves of information often require time-consuming data mining and inhibit timely advancement of health and lifestyle science. User-friendly tools for information retrieval from scientific literature such as research articles, clinical studies, and published texts have yet to be pioneered. We developed a straightforward, time effective structured knowledge repository and scientific workflow capable of providing the foundation for accelerated creation of health-focused ontologies. This semi-automated workflow enables conversion of textual annotations from scientific literature into triples (knowledge propositions in the form of semantic triples). A semantically enabled backend repository stores triples combined from many sources. Knowledge gleaned from multiple, sometimes-conflicting sources enables these triples from many sources of literature into one conceptual map with visualization of new and unexpected relationships in the form of a conceptual lattice. With the help of Protégé, these conceptual lattices convert to health-focused ontologies, which equip DSS with knowledge regarding PA and DI. Ultimately, health-focused ontologies and DSS provide patients, physicians, and researchers easy access to knowledge on health trajectories, health improvement, and individual health outcomes. By employing this semi-automated workflow and enabling concept lattice to ontology conversion, we have created an express tool for health-focused data extraction. With this system, modern medicine can embrace the idea of health promotion, rather than disease risk assessment.

DESIGN AND METHODS

We employed open source and commercial off the shelf technologies including Zotero [1], Excel [2], MySQL [3], Python [4], Cmap [5], and Protégé [6] as part of the semi-automated workflow for easy data mining and concept lattice extraction from literature. This workflow begins in Zotero's PDF viewer where human annotation takes place to highlight and note semantic triples of interest in an article as illustrated in Fig. 1. Next, the "extract annotations" tool in Zotero is used to create a .txt file, shown in Fig. 2, of the annotations made. The information in this .txt file is then transferred to Excel where a macro parses the annotations into four columns as represented by Fig. 3. The .csv file created in Excel is then imported into a table in MySQL and further parsed into a three-column table shown in Fig 4. The table in MySQL is exported as a .txt file and imported as "Propositions to text" in Cmap, creating a concept map, part of which can be seen in Fig. 5. Finally, the concept maps obtained from such articles can be exported as .cxl files, reformatted to .owl files, and imported into Protégé for ontology creation. As an example, we utilized this semi-automated workflow to extract information from "Potential adverse cardiovascular effects from excessive endurance exercise" by O'Keefe et al. and create a conceptual

lattice about the effects of PAs with varied types, intensities, durations and frequencies on CVHF [7]. A total of 177 unique concepts, 49 linking phrases, and 156 propositions were compiled from the article. These concepts are linked to concepts in other maps created from ontologies, for example *The Foundational Model of Anatomy Ontology* [8].

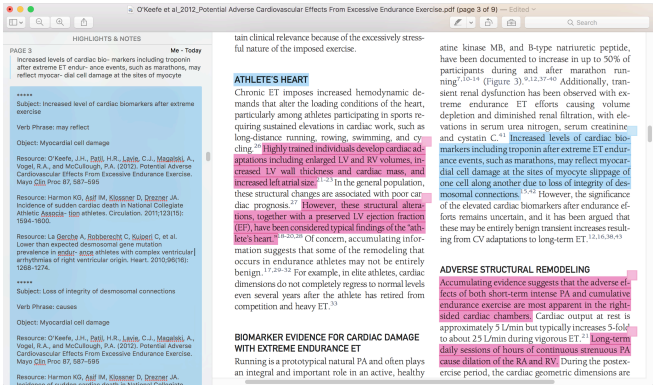


Fig. 1. Article annotations in Zotero's PDF viewer.

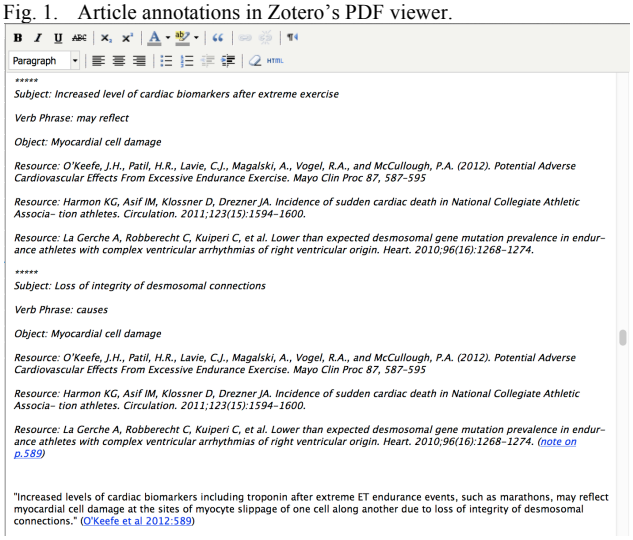
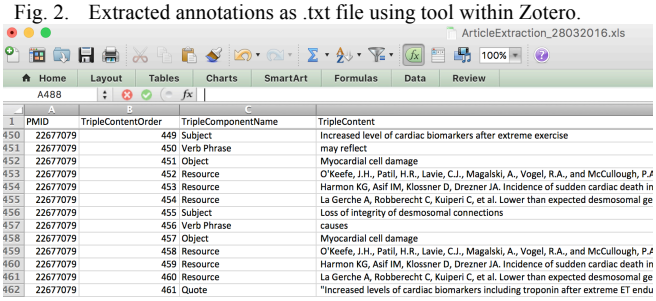
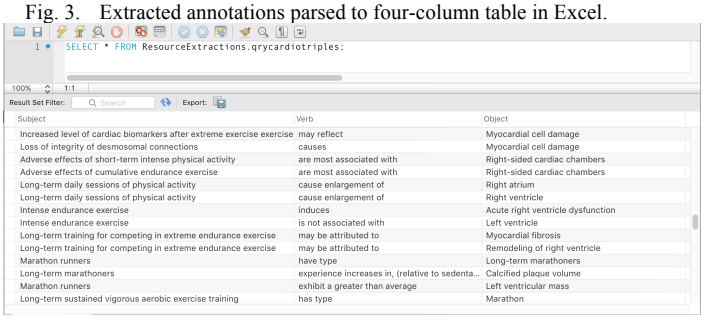


Fig. 2. Extracted annotations as .txt file using tool within Zotero.



	A	B	C	D
1	PMID	TripleContentOrder	TripleComponentName	TripleContent
450	22677079	449	Subject	Increased level of cardiac biomarkers after extreme exercise
451	22677079	450	Verb Phrase	may reflect
452	22677079	451	Object	Myocardial cell damage
453	22677079	452	Resource	O'Keefe, J.H., Patil, H.R., Lavie, C.J., Magalski, A., Vogel, R.A., and McCullough, P.A
454	22677079	453	Resource	Harmon KG, Asif IM, Klossner D, Drezner JA. Incidence of sudden cardiac death in
455	22677079	454	Resource	La Gerche A, Robberecht C, Kuiperi C, et al. Lower than expected desmosomal ge
456	22677079	455	Subject	Loss of integrity of desmosomal connections
457	22677079	456	Verb Phrase	causes
458	22677079	457	Object	Myocardial cell damage
459	22677079	458	Resource	O'Keefe, J.H., Patil, H.R., Lavie, C.J., Magalski, A., Vogel, R.A., and McCullough, P.A
460	22677079	459	Resource	Harmon KG, Asif IM, Klossner D, Drezner JA. Incidence of sudden cardiac death in
461	22677079	460	Resource	La Gerche A, Robberecht C, Kuiperi C, et al. Lower than expected desmosomal ge
462	22677079	461	Quote	"Increased levels of cardiac biomarkers including troponin after extreme ET endu

Fig. 3. Extracted annotations parsed to four-column table in Excel.



Subject	Verb	Object
Increased level of cardiac biomarkers after extreme exercise exercise	may reflect	Myocardial cell damage
Loss of integrity of desmosomal connections	causes	Myocardial cell damage
Adverse effects of short-term intense physical activity	are most associated with	Right-sided cardiac chambers
Adverse effects of cumulative endurance exercise	are most associated with	Right-sided cardiac chambers
Long-term daily sessions of physical activity	cause enlargement of	Right atrium
Long-term daily sessions of physical activity	cause enlargement of	Right ventricle
Intense endurance exercise	induces	Acute right ventricle dysfunction
Intense endurance exercise	is not associated with	Left ventricle
Long-term training for competing in extreme endurance exercise	may be attributed to	Myocardial fibrosis
Long-term training for competing in extreme endurance exercise	may be attributed to	Remodeling of right ventricle
Marathon runners	have type	Long-term marathoners
Long-term marathoners	experience increases in, (relative to sedenta...	Calcified plaque volume
Marathon runners	exhibit a greater than average	Left ventricular mass
Long-term sustained vigorous aerobic exercise training	has type	Marathon

Fig. 4. Extracted annotations parsed to three-column table in MySQL.

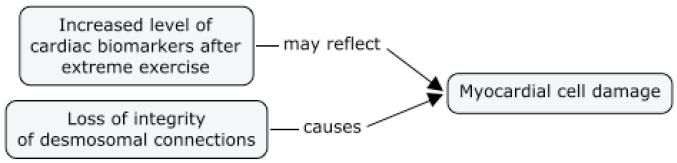


Fig. 5. Two concepts from article in CMAP.

Sustainability plans for the ontology will be developed once we receive initial feedback from the community about how paths forward for integration with related ontologies. We have not yet tested this initial ontology.

Conclusions and Further Research

We have created a prototype platform for semi-automated concept lattice generation from data mining that is easy to use, integrates information, and creates visualization for a knowledge network. It enables health professionals in preventing health problems before they start, bringing an enormous change to the medical industry. Immediate implications of this workflow are the creation of a health-focused ontology for individuals who engage in vigorous exercise and their physicians who may use it as a teaching tool. The health-focused ontology built on PA can be combined with the creation of other health-related ontologies related to PA, DI, and other health improvement methods, as part of a multi-ontology framework to accelerate the development of health promotion [9]. Correlational relationships discovered from integration of multiple ontologies will provide foundations for more research on health promotion. Further automation of this semi-automated workflow will make health-focused ontology creation even faster and more easy to use. Part of this automation process will employ development of add-on functions within Zotero, eliminating the use of Excel and extracting concepts directly into the database. Additional steps would include crowd-sourcing information by enabling this tool to communicate through web services into cross-disciplinary conceptual lattices. The goal is to develop an environment where, with minimal oversight, one can move from textual annotations into map creation easily. Ultimately, this will lay the foundation for building a large repository of structured knowledge related to PA and provide a model for mapping other human behaviors to individual health outcomes. However, in working with this prototype semi-automated workflow, errors involving imprecise language and varying tense highlight the need for detailed inspection and refinement of annotations. These errors emphasize areas of ambiguous jargon used in health, which need to be explicitly characterized. Such manual inspections take considerable time and underscore the need for semi-automated concept/linking phrase suggestion mechanisms. Despite its errors, this prototype semi-automated workflow serves as the solution for the dire necessity of a fast, accessible, and comprehensible system for improving current knowledge and information about health promotion in medicine.

REFERENCES

[1] Roy Rosenweig Center for History and New Media. *Zotero*. Retrieved from https://www.zotero.org

[2] Microsoft Office (2016). *Microsoft Excel.* Retrieved from https://www.microsoftstore.com/store/msusa/en_US/pdp/productID.323021400?s_kwcid=AL!4249!3!105984118253!e!!g!!excel&WT.mc_id=pointitsem+Google+Adwords+5+-+Excel+2016&invsrc=search&ef_id=UsDsQwAAAWWySTEQ:20160701200844:s

[3] Oracle Corporation and/or its affliliates (2016). *MySQL*. Retrieved from https://www.mysql.com

[4] Python Software Foundation (2016). *Python*. Retrieved from https://www.python.org

[5] Florida Institute for Human & Machine Cognition (2014). *Cmap*. Retrieved from http://cmap.ihmc.us

[6] Stanford Center for Biomedical Informatics Research (2016). *Protégé.* Retrieved from http://protege.stanford.edu

[7] J. O'Keefe et al., "Potential adverse cardiovascular effects from excessive endurance exercise." Mayo Clinic Proceedings 87.6, pp. 587–595, 2012.

[8] C. Rosse, J. Mejino, "A reference ontology for biomedical informatics: the foundational model of anatomy." Journal of Biomedical Informatics 36.6 pp. 478–500, 2003.

[9] M. Lange, D. Lemay, J. German, "A multi-ontology framework to guide agriculture and food towards diet and health." Journal of the Science of Food and Agriculture 87 pp. 1427-1434, 2007