

IBM Research Report

A Unified Approach for Social-Medical Discovery

Haggai Roitman, Yossi Mesika, Yevgenia Tsimerman, Sivan Yogev
IBM Research Division
Haifa Research Laboratory
Mt. Carmel 31905
Haifa, Israel



Research Division

Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

A Unified Approach for Social-Medical Discovery

Haggai ROITMAN^{a,1}, Yossi MESSIKA^a Yevgenia TSIMERMAN^a Sivan YOGEV^a

^a IBM Research, Haifa 31905, Israel

Abstract. In this paper we describe a novel social-medical discovery solution, based on an idea of *social and medical data unification*. Built on foundations of exploratory search technologies, the proposed discovery solution is better tailored for the social-medical discovery task. We then describe its implementation within the *IBM BlueMedics* system and discuss a sample usecase which demonstrates several new social-medical discovery opportunities.

Keywords. social-medical discovery, entity-relationship graph, IBM BlueMedics

1. Introduction

In recent years, *social-media (web 2.0)* has become one of the main driving forces on the web. Unlike traditional semantic-web technologies, which mainly focus on efficient interoperable data exchange among computers, social media technologies focus on online collaboration and knowledge sharing among people. Nowadays, the healthcare domain exhibits a similar shift towards the adaptation of social web technologies. New *social-medical* flavored web services are now emerging, *empowering patients* to take more active part in managing their health wellbeing. Such services offer new set of tools for sharing personal social and medical data, and sharing experiences or expertise about various health-related topics through social collaboration between patients, physicians, and various healthcare service providers.

In this line of services, online services such as *Google Health*² and *Microsoft HealthVault*³ now allow patients to share their personal health records (PHR); this compared to traditional EMR systems that prohibit patients from accessing their own medical records. Depending on patients' privacy preferences, PHR data may be publicly (or partially) shared, offering new discovery opportunities. For example, personalized medical content recommendations may be delivered to patients based on their PHR data [6]. Furthermore, several online social-medical community services such as *Patients-Like-Me*⁴ and *Cure-Together*⁵ allow patients to discover other patients who share similar medical characteristics, such as similar disorders or symptoms. By joining to a medical community, patients

¹Corresponding Author: Haggai Roitman, IBM Research, Haifa 31905, Israel; E-mail: haggai@il.ibm.com.

²<http://www.google.com/health/>

³<http://www.healthvault.com/>

⁴<http://www.patientslikeme.com/>

⁵<http://www.curetogether.com/>

may get additional medical (and even mental) support, which leverage the community's power to discover new possible treatment plans, clinical trials, expert physicians, etc. Finally, online services such as *Med-Help*⁶ and *Drugs.com*⁷ provide access to rich medical knowledge gathered from various medical knowledge resources (e.g., HCLS Linked Open Drug Data (LODD)). Such resources can be used by users who seek drug-related information, wish to find expert advice, or find evidence for various health related topics.

Despite the increasing amounts of social data fused together with rich medical data, there still remains a great challenge of how to fully utilize this new combination for purposes of efficient *social-medical discovery*. Existing social media discovery solutions use relatively simple data models that record relationships between people and their associations with unstructured (text) documents [2]. Therefore, existing social data models are not well suited for handling medical data, which is usually structured in its nature, semantically rich (e.g., defined over some medical terminologies such as SNOMED-CT, UMLS, ICD-10, etc), standard-based (e.g., HL7 RIM), etc. On the other hand, existing social-medical solutions utilize only simplified data models and provide limited discovery capabilities that merely exploit the *social-medical dataspace*. For example, social community services currently provide very simple query interfaces for exploring their social-medical community data, spanning from simple keyword search to very limited category-based search over several medical facets such as symptoms or demographic data. As another example, personalized medical recommendation systems that utilize PHR data commonly ignore social data. Furthermore, patients, and even more expert users such as physicians, usually find it hard to explore data that they are not familiar with its structure, terminology, query language, etc [3]; hence, a more *exploratory* solution [5] is desired which can gradually guide patients within the social-medical dataspace. Such data exploration should be backed up with as much *evidence* as possible, yet very *intuitive* even for non-expert users such as patients usually are.

Aiming at fulfilling the gaps, in this paper we describe a novel social-medical discovery solution, based on an idea of *social and medical data unification*. Built on foundations of exploratory search technologies, the proposed discovery solution is better tailored for the social-medical discovery task. In the rest of this paper we describe our solution, its fundamentals and discovery capabilities. We then describe its implementation within the *IBM BlueMedics* system.

2. Methods

We now present a novel model for social-medical discovery. Built on foundations of conceptual modeling, social data and medical data are fused together using a uniform representation in the form of a rich entity-relationship (ER) data graph. In turn, social discovery can be augmented with medical discovery and vice-versa. This allows to explore new facts about social and medical entities through various paths within the ER graph. For example, we may discover similar patients not only based on direct patient similarity, but also based on their relationships with other similar social or medical entities, e.g., similar medications, allergies, family bonds, treating physicians, etc.

⁶<http://www.medhelp.org/>

⁷<http://www.drugs.com/>

Social and medical facts known to exist are modeled by entities and their relationships⁸. Such facts can be gathered by observing and collecting data from various data sources, such as the ones that were mentioned in Section 1. Each fact is accompanied with an evidence link which trace its source origin, e.g., a fact about an adverse drug-reaction between two drugs may be linked to its FDA alert page or knowledge from *DrugBank*⁹. Knowledge may be gathered either by manual or semi-automatic means, e.g., by using entity extraction, resolution and disambiguation techniques [4]. Social entities include among others, patients, physicians, or even “virtual entities” such as various health service providers (e.g., hospitals). Medical entities include among others, medications, allergies, immunizations, symptoms, genetic variations, etc. Each entity may have a rich set of attributes describing its properties. For example, a patient is represented as a single entity in the graph together with its socio-demographic attributes such as gender, age, location, etc. As another example, each medication is represented as an entity with attributes such as its generic or brand name (code), substance name, etc. Both social and medical entities may have relationships with other social or medical entities. For example, a patient entity may be related to some consumed medication; a medication entity may be related to some drug-interacting medication entity; a patient entity may be related to his treating physician entity, etc.

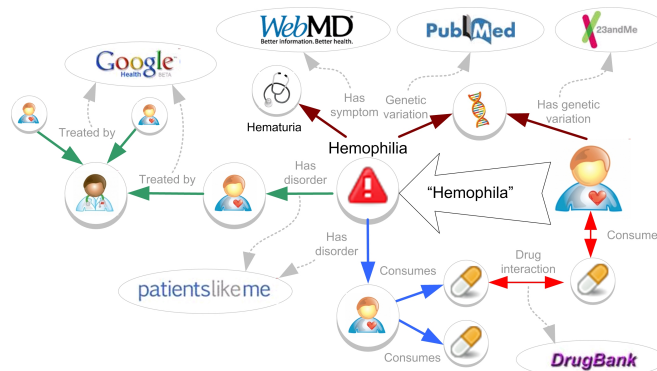
Using such a discovery model allows to support various types of exploratory-driven queries over the social-medical data graph. This includes rich keyword-based queries that can also be mixed with more structured query predicates, allowing to express very complex information needs. For example, users can submit a query like “Hemophilia AND Patient.age:[40 TO *)” to discover all patients whose age is above 40 and are related to Hemophilia related topics. Furthermore, the discovery model supports rich faceted-search and data lineage capabilities [5] that allow interactive exploration of the social-medical dataspace. For that, we extended the basic faceted-search model of [1] and implemented a novel inverted index structure that enables to index and retrieve both text and structured data formats and supports an OLAP-like complex faceted search over rich entity-relationship data. Using the faceted-search user interface, users may start their search based on some initial information need. Search result includes a list of social or medical entities or both, uniformly ranked by their relevance to the user’s query. Each entity is further accompanied with relationship links that allow the user to flexibly explore the sub-graph induced from that entity. In addition, facets about various entities in the result set (e.g., patient age or gender distributions) further allow the user to quickly filter out entities according to facets of interest and explore the graph projected by those facets. Such social-medical data exploration may be highly useful for patients who wish to explore possible treatment plans by following the medication links of some patient returned as result to their query, or for physicians and researchers who wish to discover new interesting patterns in the social-medical dataspace. We discuss more example usages in the next section.

⁸Our discovery model makes no assumption on the exact ER data representation. Though, a natural way to represent ER data is using a relational data model or in RDF.

⁹<http://www.drugbank.ca/>



(a) IBM BlueMedics social-medical discovery (SMD) user interface



(b) Example social-medical sub-graph for the query "Hemophilia"

Figure 1. Example social-medical discovery use case implemented within IBM BlueMedics

3. Results

We have implemented the proposed social-medical discovery solution within the *IBM BlueMedics* system. *IBM BlueMedics* is a novel clinical decision support system (CDSS) developed in collaboration between three IBM labs and the GIL hospital in Korea. *IBM BlueMedics* empowers the patients and helps to increase patient safety by assisting patients and their medical providers with daily medical decision-making. One of the main services in *IBM BlueMedics* is the social-medical discovery (SMD) service. SMD serves various queries, submitted by patients, physicians, and researchers, which explore its social-medical dataspace. *IBM BlueMedics* social-medical dataspace is formed by integrating social and medical data stored within *IBM BlueMedics* sub-systems together with data it gathers from public social-medical sources.

Figure 1(a) depicts SMD's main user interface with an example discovery usecase. In this example, two patients were returned as a result to an initial query "Hemophilia" submitted by the user, who later on followed the "Related patients" link to discover

relevant patients. We can also observe that SMD provides several facets related to those patients, and for each patient, the user may further follow several relationship links to explore that patient's social-medical sub-graph.

Finally, to illustrate additional discovery options, Figure 1(b) further depicts a possible social-medical sub-graph that may be explored by a patient user searching for Hemophilia-related information. By following the links to Hemophilia-related patients (e.g., based on information gathered from *Patients-Like-Me* or *Google Health*), a patient searcher can discover new possible treatments, e.g., other medications consumed by those patients or physicians who treat those patients, and whom the patient may contact for her own benefit. Furthermore, the searcher may discover Hemophilia-related symptoms (e.g., Hematuria) gathered from *WebMD.com*¹⁰, or discover related genetic variations gathered from *PubMed*¹¹, etc. Using the searcher's own medical profile, she can also discover possible lineage paths between her genetic profile (e.g., from *23AndMe*¹²) and Hemophilia. The patient can also detect whether a new medication, which she just discovered through some related patient, has a potential interaction (e.g., based on *Drug-Bank* gathered knowledge) with any of the medications she currently consumes.

4. Conclusions

In this paper, we suggested a novel unified social-medical discovery model based on ideas of conceptual modeling. We described the model and discussed its capabilities. We then shortly described its implementation within *IBM BlueMedics* and described several new discovery opportunities based on a sample usecase. As future work, we plan to utilize the new social-medical discovery model for purposes of expert search and patient community analysis. We also plan to extend the basic model with contextual information such as temporal data and medical context. This in turn, may help to detect hidden trends within the social-medical dataspace or be used for search personalization.

References

- [1] Ori Ben-Yitzhak, Nadav Golbandi, Nadav Har'El, Ronny Lempel, Andreas Neumann, Shila Ofek-Koifman, Dafna Sheinwald, Eugene Shekita, Benjamin Sznajder, and Sivan Yogev. Beyond basic faceted search. In *Proceedings of the international conference on Web search and web data mining, WSDM '08*, pages 33–44, New York, NY, USA, 2008. ACM.
- [2] Peter J. Carrington, John Scott, and Stanley Wasserman. *Models and Methods in Social Network Analysis (Structural Analysis in the Social Sciences)*. Cambridge University Press, February 2005.
- [3] R. J. W. Cline. Consumer health information seeking on the Internet: the state of the art. *Health Education Research*, 16(6):671–692, 2001.
- [4] Brizan D. Guy and Tansel U. Abdullah. A survey of entity resolution and record lineage methodologies. *Communications of the IIMA*, 6:41–50, 2006.
- [5] Gary Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49:41–46, April 2006.
- [6] Haggai Roitman, Yossi Messika, Yevgenia Tsimmerman, and Yonatan Maman. Increasing patient safety using explanation-driven personalized content recommendation. In *Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10*, pages 430–434, New York, NY, USA, 2010. ACM.

¹⁰<http://symptoms.webmd.com/>

¹¹<http://www.ncbi.nlm.nih.gov/pubmed>

¹²<https://www.23andme.com/>