SBRS: Bridging the Gap between Biomedical Research and Clinical Practice

Santiago Timón-Reina, Rafael Martínez-Tomás, Mariano Rincón-Zamorano, Tomás García-Sáiz, Estela Díaz-López, and R.M. Molina-Ruíz

UNED, Hospital Clínico San Carlos

Abstract. The field of Biomedical research is currently one with the greatest social impact and publication volume, providing continuous advances and results which should, to a great extent, reach the general clinical practice. Similarly, direct clinical experience may offer experimental results and conclusions which may lead, guide and foster new investigations. However, this interaction between research and clinical practice is yet too far from being optimal. On one side, research results are published without standardization, suffering terminological issues, which prevent its automatic handling and great scale information treatment/management. On the other, for the practitioner, the task of reviewing papers, bibliography, experimental results, etc. in order to keep updated his everyday clinical practice, is very time consuming, causing not to be done continuously.

The implantation of Information Technologies in the biomedical research field has developed numerous search and bibliographic management resources, existing a current trend towards building and publishing open access terminologies, ontological knowledge models and big datasets with biomedical content. All together, beside Semantic Web technologies, methodologies and Linked Open Data and AI techniques, conforms a technological framework which gives the opportunity to bridge the gap between research and clinical practice to support the physician in evidence based decision making.

In this work, as a starting point to the final aim of linking research and clinical practice, we describe a Semantic Bibliographical Recommender System (SBRS) based on patient profile integrated with electronic health record (EHR) which, without closing the loop, offers to the medical professional the latest and most significant experimental evidences related to his concrete case study. The system's functionality and utility is exemplified through real life psychiatric cases, assisted by an expert psychiatrist.

1 Introduction

The field of Biomedical research is one of most social impacting and publishing volume fields. This fact is easily observable through some bibliographic search systems indexation evolution. To exemplify, we have used "Medline Trend" tool [1] to obtain the evolution of the number of PubMed indexed papers per year,

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 191–200, 2013.

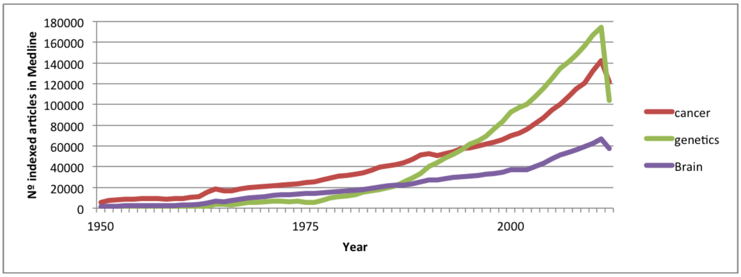

Fig. 1. Publication volume evolution

performing three search queries by general and popular terms in the Biomedical context: cancer, genetics and brain (for the term neuroscience is relatively new). Figure 1 shows this evolution.

The discoveries made by this research field usually have an application in clinical practice, either directly (clinical research) or indirectly (basic science), updating diagnosis, protocols, etc. In the same way, clinical experience may give conclusions and experimental results which impact directly on current research, guiding and fostering new investigations. A very clear example is novel drug tracking in clinical practice, which have its own standards and protocols [2].

However, the interaction between Biomedical research and clinical practice is still far from being efficient. Despite the final aim of research is finding ways of application, divulgation means are specific and closed: few search engines and online journals with few application interfaces. This situation requires a proactive attitude from interested people and agencies looking for published articles. Moreover, research papers lack of the standardization required to allow automatic treatment beyond its indexation. All together results in a very time consuming task, making it difficult for the doctor to remain updated in his daily clinical practice. It is therefore necessary a search for means and specific strategies to ease and encourage implementation of research based recommendations and to ensure changes in practice [3].

Fortunately, with the gradual (and faster everyday) implantation of Information Technologies and Artificial Intelligence techniques, tools have emerged to help the dissemination, search and access to literature, such as PubMed[1]. Likewise, working groups focused on formalizing the terminology along different biomedical knowledge domains started to appear, developing terminologies such as SNOMED CT [4] or MeSH[2] and, a step beyond the terminologies, comprehensive ontology based knowledge models like Gene Ontology (GO) [5] or the Foundational Model of Anatomy (FMA) [6].

[1] http://www.ncbi.nlm.nih.gov/pubmed/
[2] http://www.nlm.nih.gov/mesh/

Following standards and methodologies of the Semantic Web[3] (OWL 2[4], RDF[5], SWRL[6] and SPARQL[7]), these models are enabling the build of great linked, opened and semantic datasets, following Linked Open Data[8] recommendations.

Altogether, we have a) bibliographic search engines focused in Biomedical research, b) controlled terminologies and knowledge models over various biomedical knowledge domains and c) great datasets, giving the opportunity to create, on the one hand, standards for structuring and sharing Biomedical research articles and, on the other hand, tools for exploiting those standards to ease the access to research content for the practitioners.

Given the number and diversity of research domains and the high level of domain-specific expertise that is needed for such an undertaking, it is only reasonable that each discipline take responsibility for developing its research abstraction scheme. However, given the Semantic standards, it will be desirable that there would be an abstraction framework to easily manage disparate research works.

In the field of neuroimaging, the BrainMap database project has been one of the first efforts aiming at the standardization problem [7]. It defines a formal metadata coding scheme to describe the content of functional neuroimaging research, allowing to search across coordinate and brain report locations.

As we have stated, the automatic bibliographic management is a very ambitious goal, requiring the approach and resolution of various problems. The first of them is the low degree of standardization in natural language paper structure and the need for a digital content description beyond terminological indexation.

Despite the necessary degree of standardization is not yet available, taking this technological framework as a starting point, it is possible to design and build systems which may take advantage of available technology. In this work we have developed a prototype of a Semantic Bibliographic Recommender System (SBRS) for the field of neuroimaging which, based on a Patient Profile, generates bibliographic references related to patients characteristics. In order to achieve this goal, we have mapped and extended various bio-ontological resources, generating an ontological model to represent patients and neuroimaging articles. With a system like the one proposed, a practitioner could receive research feedback while seeing patients.

2 Methods

In Web context, user experience has been notably improved in the last years with the introduction and growth of Recommendation Systems. These systems

[3] `http://www.w3.org/standards/semanticweb/`
[4] `http://www.w3.org/TR/2012/REC-owl2-overview-20121211/`
[5] `http://www.w3.org/standards/techs/rdf#w3c_all`
[6] `http://www.w3.org/Submission/SWRL/`
[7] `http://www.w3.org/standards/techs/sparql#w3c_all`
[8] `http://www.w3.org/standards/semanticweb/data`

proactively suggest items which may be of particular interest to the user, based on his behavior and/or preferences[8].

Our initial proposal to improve biomedical research integration within everyday clinical practice is a Semantic Bibliographic Recommender System (SBRS), which follows a Content Based approach [9]. In this kind of systems, items are described based on its features. Since these items are research papers and we are following a semantic approach, these features are concepts from the knowledge model.

In order to build recommendations, the system needs some kind of input or query and a known user profile. This profile usually consists in two types of information:

1. A model of the user's preferences. One common representation is a function which for any item predicts the likelihood that the user is interested in that item. In our case is a semantic description of the patient.
2. A history of the user's interactions with the system. This history helps to improve system's performance by learning and adjusting function's components.

Unlike common Recommender Systems, where the user profile is built on one entity (the user), our scenario requires to split the information between the profile which provides the features of interest and the interaction history. The patient characteristics conform the features of interest, building the Patient Profile (PP), and the interaction history would be built on physician's item choice. At this moment, the system stores user feedback and interaction, but the learning process is not yet implemented.

2.1 Patient and Item Modeling

In order to build such a Recommendation System, we need to define an item description model describing the items the system will have to serve. In our case, the items are neuroimaging research papers (NRP). As we have already seen, Brainmap's metadata codification scheme gives us a good starting point but, since we are looking for semantic interoperability between different domains (anatomy, psychology, genetics, etc.), we need to build a semantic model of this scheme i.e, an ontology.

Fortunately there already exists an ontology covering part of Braimap's coding scheme: the Cognitive Paradigm Ontology [10]. This project intends to formalize, starting from Brainmap Scheme, certain characteristics of the cognitive paradigms used in the fMRI and PET literature. However, it does not fully map concepts with other domain ontologies (such as FMA) and neither every Brainmap metadata is represented. Hence, in order to cover our needs, we have needed to add some extensions, generating the extended CogPOe.

Primarily, the extensions we need are Patient/Subject related. Since our system makes use of patient profiles, the needs can be covered by aligning Patient and EHR oriented ontologies. The Computer-based Patient Ontology (CPR) [11]

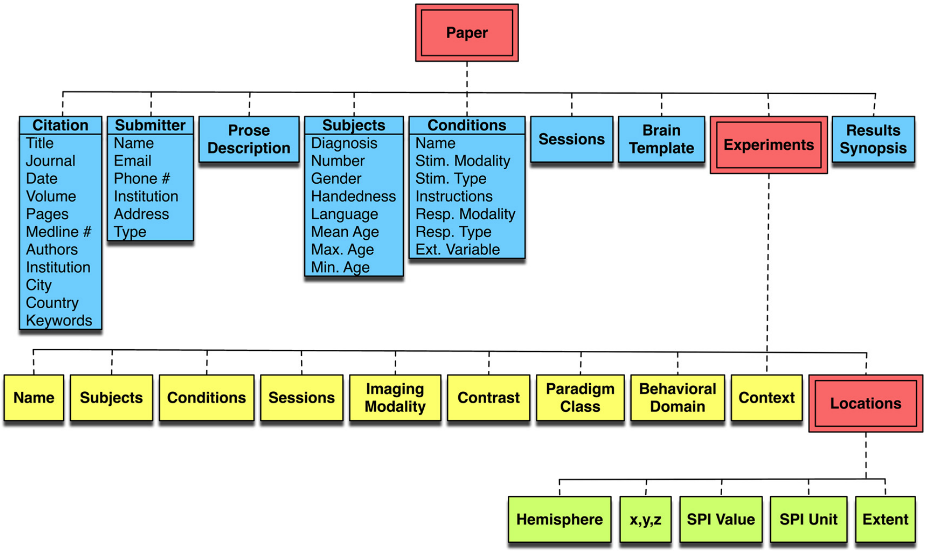

Fig. 2. Brainmap meta-data codification scheme

attempts to define a minimal set of terms that provide grounded, ontologically commitment for the representations shared between many of the healthcare information (such as HL7 RIM)[9], process and terminological models via the use of foundational ontologies.

Briefly, the alignment has been done by referencing CPR concepts from CogPOe. The most important mapping is the relation between Brainmap's Subjects concept, because it defines the subjects involved in an experiment concept and will, indirectly, carry many relevant information. It is represented with cpr:Patient concept. Related patient features (properties), such as diagnostic, gender, medication, etc are also obtained from CPR.

2.2 Building the Patient Profile

As we already said, the system will be based on patient features, which means that it must be build from some sort of EHR data.

But user related data in the context of EHR is a delicate issue because of the existence of segmented EHR systems managing disparate patient representations and, more importantly, the use of various standards among different regulatory agencies.

The best way to deal with this problem is enabling semantic interoperability between EHR systems and regulatory bodies , by implementing a Semantic Mediation System capable of automatically map equivalent fields or concepts between different standards [12]. This is the solution raised by the SALUS project [13,14].

[9] `http://www.hl7.org/implement/standards/rim.cfm`

The Patient Profile building process is designed following the same principles, creating a semantic representation of patient profile from EHR's input.

Therefore, we need to choose an EHR standard which will serve as the system's input. There exist multiple EHR standards so, based on its widespread implementation, we chose HL7's (Health Level 7)[10] Clinical Document Architecture (CDA)[11] were chosen. Briefly, this standard enables documents to be expressed both as free text and coded format, using SNOMED CT as its terminology.

As shown in figure 3, using CDA and SNOMED codes is possible to map physical exploration document concepts to CPR.

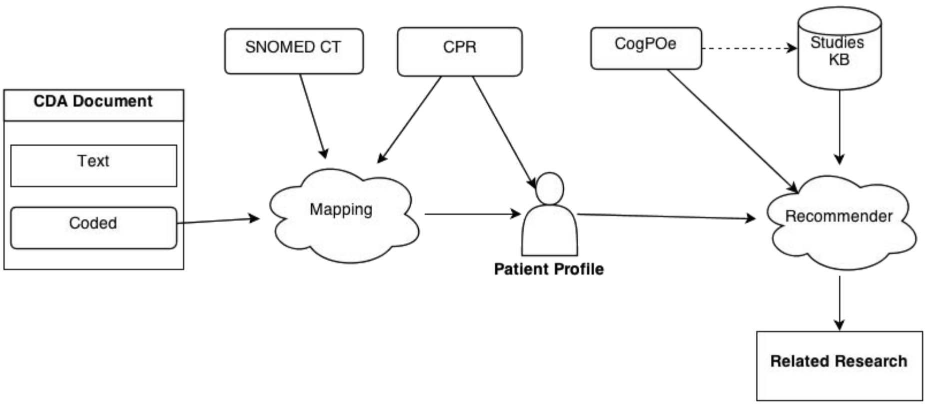

Fig. 3. System overview

2.3 NPR Selection

Both the Patient Profile and item descriptions are instances of our patient centered ontological model, so, in order to evaluate the relevance of a given paper, we have to compare the patient's features with research subject's features. This is achieved with a similarity function which quantifies the similarity between two given instances.

Looking into the ontology mapping literature, we find many similarity measuring methods: hierarchical, graph-based, information theory-based, etc. Since we are looking into instances rather than concepts (i.e., classes) we need to explore the way these instances express its meaning: the implicit labeled graph structure of semantics. We have implemented the similarity function by traversing the ontology instances [15].

According to this method, in order to evaluate the similarity of a pair of given instances, $InstA$ and $InstB$, we need to look into the set of properties which connects them to other elements. Let this instances have properties

[10] `http://www.hl7.org/`

[11] `http://hl7book.net/index.php?title=CDA`

$A = \{property_1, \ldots, property_n\}$ and $B = \{property_1, \ldots, property_m\}$ respectively. For elements connected by a common property i, the similarity measure is computed by the $PropertySM_i$ function.

Then, the similarity measure for two instances is computed as the sum of similarity measures obtained for each property:

$$similarity(InstA, InstB) = \frac{\sum_{i=1}^{|A \cap B|} PropertySM_i(elemtA, elementB)}{|A \cup B|}$$

where $elementA$ and $elementB$ are elements (instances or literals) connected to the evaluated instances $InstA$ and $InstB$ by the i^{th} property respectively. This function gives a similarity measure from the range $[0, 1]$.

At this point, the similarity function sets the same weight factor for each property. This factor is needed to personalize and to determine the importance of a given property, which is context dependent. For example, in our problem, the property hasDisease may have more weight than hasName, since the diagnosis is more important than the name to compute the Patient Profile similarity.

3 Use Case

In order to illustrate the viability of our solution, we have tested our system reproducing, under the proposed structure, with a real clinical case with the collaboration and feedback from an expert psychiatrist focused on Eating Disorder (ED) research.

Classifications of ED (DSM-IV and ICD10) are still mainly focused on the preoccupation for body weight and distortion of body shape, which do not really seem to tell enough about these disorders, specially regarding treatment. More recently, affect dysregulation in ED has been emphasized [16,17], and there is quite a lot of evidence that the emotional awareness and emotion regulation are affected in ED [18]. Most of these ideas were supported by studies that emphasized the relationship between alexithymia and emotional awareness in anorexia. Recent studies are focused in the neurocircuits behind these processes and even recent efforts to reclassify these disorders have suggested classifications of ED within personality subtypes in a: 1) dysregulated/undercontrolled pattern, 2) a constricted/overcontrolled pattern, 3) high-functioning/perfectionistic pattern [19].

Despite the ongoing research efforts, many of these suggestions and discoveries are unnoticed in everyday clinical practice.

As we already stated, at this moment, there is not available any knowledge/semantic database with full-structured research papers. For this reason, we have needed to build a knowledge base using Virtuoso-Opensource[12] triple store where a set of 113 papers, following our ontological model (CogPOe), have been stored. With this dataset we have been able to test the system with the clinical case.

---

[12] http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/

This clinical case consisted in a patient diagnosed with Anorexia Nervosa. The psychopathological exploration identifies that the patient is calmed with collaborative attitude and fluid and coherent speech., but with hypercontrol and obsessive personality traits.

We have tested the system with an example of ED coded in a CDA compliant document serving as the input.From this document, the system maps and builds the PP to compute the set of nearest bibliographic recommendations. The output of this very first query was satisfactory, since 8 of 10 from retrieved NPRs were considered as related to the clinical case, being the 80% of recommendations. Because of the limited dataset size we still have not been able to plan a rigorous performance evaluation (like Precision and Recall), being the dataset population as one of the future works. However, the expert remarked that having this information available from concepts related to the spectrum of these disorders, e.g. ALEXITHYMIA, one of the main characteristics described in AN, would be very useful when trying to decide the best therapy for these patients. For example, patients more focused on emotional regulation should incise in mindfulness therapy. So, this practical example shows us the utility of these integrated tools for the practitioner.

4 Conclusions and Future Work

In this work we have noted a very known problem: the gap between biomedical research and clinical practice. We also have highlighted that this gap is getting narrower as technology is being implanted in Health Care Systems. However, more technology means different systems interoperating and, many of them, overlapping, arising the need for a greater standardization.

As a sample of this standardization, we have created a semantic model starting from Brainmap's metadata coding scheme, reusing, mapping and extending disparate domain ontologies (CogPO and CPR). With this representation of neuroimaging papers, we have been able to propose a Semantic Bibliographic Recommender System to help the practitioner in the bibliographic gathering task. Planning the integration with EHR systems, we have design our system based on the widely used HL7 CDA standard, creating a mapping system to obtain a Patient Profile as the Recommender input.

To test our proposal, we have worked with an expert psychiatrist, receiving advice and feedback. The tests results are promising, encouraging us to keep populating the dataset, refine and scale the system to greater goals.

The concept and the architecture of the Web is naturally evolving to a distributed service oriented scenario, emerging applications which implement interfaces to allow third party applications to interact with them. We think bioinformatic applications should be compliant with this philosophy, therefore, our system is built as a RESTful service, allowing any application to interact with it.

The semantic Patient Profile opens the door to integrate large amounts of clinical data, beyond research bibliography. Linked Open Data initiative is spreading along many of the Biomedical domains and, with tools like Bioportal and the Datahub[13], it is possible to easily locate, use and share existing biomedical data.

Developing this integration between EHR systems and federated biomedical datasets is no longer a wish, but a fact with on going efforts like the Clinica Mayo's [20].

In the future, we look forward to integrate drug and genetic datasets to automatically enrich the Patient Profile, giving the chance to offer the most complete information for the practitioner.

References

1. Corlan, A.D.: Medline trend: automated yearly statistics of PubMed results for any query (2004), http://dan.corlan.net/medline-trend.html
2. C. for medicinal products for human use (CHMP), ICH guideline E2B (R3) Electronic transmission of individual case safety reports (ICSRs) - implementation guide - data elements and message specification (2011), http://www.emea.europa.eu/docs/enGB/documentlibrary/Scientificguideline/2009/09/WC500002767.pdf
3. Bero, L.A., Grilli, R., Grimshaw, J.M., Harvey, E., Oxman, A.D., Thomson, M.A.: Getting research findings into practice: Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. BMJ 317(7156), 465–468 (1998), http://www.bmj.com/content/317/7156/465
4. Stearns, M., Price, C.: SNOMED clinical terms: overview of the development process and project status. In: Proceedings of the AMIA (2001), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243297/
5. Ashburner, M., Ball, C., Blake, J.: Gene Ontology: tool for the unification of biology. Nature . . . (2000), http://www.ncbi.nlm.nih.gov/pmc/articles/pmc3037419/
6. Rosse, C., Mejino, J.L.V.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. Journal of biomedical informatics 36(6), 478–500 (2003), http://www.ncbi.nlm.nih.gov/pubmed/14759820
7. Fox, P.T., Lancaster, J.L.: Opinion: Mapping context and content: the BrainMap model. Nature Reviews Neuroscience 3(4), 319–321 (2002), http://www.ncbi.nlm.nih.gov/pubmed/11967563
8. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User Modeling and User-Adapted Interaction 22(4-5), 317–355 (2012), http://www.springerlink.com/index/10.1007/s11257-011-9115-7
9. Pazzani, M., Billsus, D.: Content-based recommendation systems. In: The adaptive web, pp. 325–341 (2007), http://www.springerlink.com/index/QQ35WT68L6774261.pdf
10. Turner, J.A., Laird, A.R.: The cognitive paradigm ontology: design and application. Neuroinformatics 10(1), 57–66 (2012), http://www.ncbi.nlm.nih.gov/pubmed/21643732
11. W3C, Computer-based Patient Record Ontology (2009), http://code.google.com/p/cpr-ontology/

[13] http://datahub.io/

12. Martínez-Costa, C., de Andrade, A., Karlsson, D.: Towards the harmonization of Clinical Information and Terminologies by Formal Representation. In: EJBI, vol. 8(3) (2012), http://www.ejbi.org/img/ejbi/2012/3/Costa_en.pdf
13. Declerck, G., Hussain, S., Parès, Y., Daniel, C.: Semantic-sensitive extraction of EHR data to support adverse drug event reporting (2011), http://www.srdc.com.tr/projects/salus/docs/ICSR_Reporting_Tool_SWAT4LS_V2.pdf
14. Laleci, G., Yuksel, M., Dogac, A.: Providing Semantic Interoperability between Clinical Care and Clinical Research Domains. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society c, 1–14 (2012), http://www.ncbi.nlm.nih.gov/pubmed/23008263
15. Andrejko, A., Bieliková, M.: Comparing instances of ontological concepts for personalized recommendation in large information spaces. Computing and Informatics 28, 429–452 (2012), http://www.cai.sk/ojs/index.php/cai/article/viewArticle/45
16. Gilboa-Schechtman, E., Avnon, L.: Emotional processing in eating disorders: specific impairment or general distress related deficiency? Depression and . . . (2006), http://onlinelibrary.wiley.com/doi/10.1002/da.20163/abstract
17. Hatch, A., Madden, S., Kohn, M.: Anorexia nervosa: towards an integrative neuroscience model. European Eating . . . (2010), http://onlinelibrary.wiley.com/doi/10.1002/erv.974/abstract
18. Ortego Saenz de Cabezón, A., Espina Eizaguirre, A.: Alexithymia and its relationships with anxiety and depression in eating disorders. Personality and . . . (2004), http://www.sciencedirect.com/science/article/pii/S0191886903000990
19. Thompson-Brenner, H., Westen, D.: Personality subtypes in eating disorders: validation of a classification in a naturalistic sample. The British Journal of Psychiatry (2005), http://bjp.rcpsych.org/content/186/6/516.short
20. Pathak, J., Kiefer, R., Chute, C.: Applying linked data principles to represent patient's electronic health records at Mayo clinic: a case report. In: . . . Symposium on International Health . . . , pp. 455–464 (2012), http://dl.acm.org/citation.cfm?id=2110415