EVALUATION METRICS FOR USER-CENTERED RANKING OF CONTENT IN METADLs

James Ford, Fillia Makedon, Li Shen, Tilmann Steinberg
The Dartmouth Experimental Visualization Laboratory (DEVLAB)
Department of Computer Science[1]

Andrew Saykin, Heather Wishart
Brain Imaging Laboratory, Dartmouth Medical School

Dartmouth College
Hanover, NH 03755, USA

I. INTRODUCTION

MetaDLs are digital libraries (DLs) designed to store and manage the access of metadata rather than original data [1]. They are particularly suited for data that is sensitive in nature (possibly proprietary), complex in structure, of high value (or production cost), and highly heterogeneous. The authors describe such a MetaDL system architecture in [9], where the model is demonstrated in the case of human brain data that can include brain activation maps, lesions, brain structures, test methodologies, patient information and tools. This paper describes evaluation metrics for such a MetaDL in computational neuroscience research. MetaDLs are applicable to complex data outside of neuroscience as well: gene sequences in genetic mapping, satellite earth/weather patterns, etc. (Other metadata DL descriptions have been given in [5, 7].)

A MetaDL exists within a two-tier architecture that supports two endeavors: searching for data (and methods) via metadata, and sharing this data in a secure fashion. Tier 1 consists of autonomous DLs containing data, each with an interface allowing it to specify access conditions. Tier 2 systems contain data about the Tier 1 DLs and permit browsing and searching for data that are contained in Tier 1 DLs. MetaDLs use meta-data to describe data objects in a uniform, structured manner, thus deriving a parameterized description of the object. Objects are then accessed, traded, manipulated, updated, etc. securely, as the original data resides with the content owner. Besides protecting the information from malicious abuse, a key additional benefit is making heterogeneous data interoperable through a common description. MetaDLs enable a user to search, browse, manipulate, and rank information, at Tier 2 level, without needing the original. Actual data exchange or access takes place at Tier 1 level, after a “data request” has been formulated and recorded by MetaDLs [1]. MetaDLs are designed to support, document and manage user-to-user transactions on a small scale.

The BrainMap system [2] and the fMRI Data Center [3] attempt to address this situation in the area of functional imaging by consolidating results and data, respectively, in central repositories. BrainMap contains summary information for various refereed publications related to HBM, including for example reported foci of activations in different studies, the number of subjects used in each, and other similar information. Users of the system can use this data for meta-analysis of the results, for instance by grouping several studies to increase statistical power. The system does not provide or facilitate exchange of raw data, however, and this limits the range of

[1] This work has been supported in part by NSF-IDM research grant number 0083423.

possible meta-analyses. It also does not include descriptions of disease-induced structural changes. The fMRI Data Center is an NSF-sponsored site that publishes raw data related to published functional imaging articles in certain refereed journals. These studies are not focused on the Human Brain and are not comprehensive by disease. Authors of published articles are required to provide their data for public access, although no particular format or level of documentation is required. It is conceivable that these data might be used for re-analysis or for combination with data from other studies, but this is hampered by the fact that available data is not indexed by features. Other related projects include METAe [5], MUMIS [6], Dublin Core [7], and ARION [8].

In [1], the authors have outlined a new digital library approach, BrassDL (BRain Access Support System Digital Library) that addresses the limitations of BrainMap and the fMRI Data Center. BrassDL aims to support collaboration, scientific exchange and discovery by a system by representing information via metadata (number of subjects, freshness of data, location, scanner parameters, etc.). Thus, brain structures (lesions, structures) and data, experiments, methods, and subject information are all represented using a common metadata representation that is inserted into the system directly by the researcher. BrassDL provides strong incentives for participation in this pooling of information through a framework of services to the research community. Besides offering visibility, ranking, and feedback, BrassDL can support meta-analysis of metadata information and the BrassDL–Interchange, a business model for managing data interchange and tracking.

In the rest of the paper, section II introduces the BrassDL, a MetaDL of human brain metadata that integrates heterogeneous brain imaging results, methods, experimental conditions, and subject information. Section III introduces a new metric for evaluating the content of a MetaDL in terms of how it matches user demand. Section IV concludes the paper.

II. BRASSDL: A METADL OF HUMAN BRAIN METADATA

As a MetaDL, BrassDL [1] exists within a two-tier architecture. Tier 1 consists of autonomous DLs controlled by data providers (or BrassDL Partners). These DLs contains primary data, metadata and information about their access conditions. BrassDL is a Tier 2 system and contains data about the Tier 1 DLs. BrassDL manages metadata collection, retrieval and evaluation, and also manages the user-to-user data exchange in a smaller scale. BrassDL collects features and parameterizes the evaluation of the dataset, thus enabling a ranking of the results as well as of the user demand, on the basis of how good the system is to the user's request, and according to user-defined preferences.

Efficient data collection is achieved through uniform and structured representation of diverse brain data: activations, lesions, brain structures, methodologies (such as segmentation), experiments and patient information (such as age, gender, handedness, education, and other) that is minimalist in nature. The user enters the information in his own secure system and the system translates this data through a client model to a MetaDL set of parameters which it then proceeds to evaluate internally. The integration of the metadata is supported by a toolkit that recognizes the type of object to be represented, thus offering the user a limited and structured set of options.

Basic searches provide a query term, such as a keyword or a sample image, that is looked up in the original data. In contrast, *advanced search* is a type of search patterned after expert boolean searches in Web search engines, but which is specific to a MetaDL. Since a MetaDL consists of meta-data items of various sorts, most with limited range of values (such as dates of entry or

Related Features	Analysis Method	Region of Interest	Disease	Num. of Subjects	Acquisition modality	
R1	Volume ($c = 1$)	Hippocampus ($c = 1$)	Epilepsy ($c = 3$)	20 ($c = 3$)	MRI ($c = 1$)	
R2	Shape ($c = 2$)	Hippocampus-amygdala ($c = 2$)	Schizophrenia ($c = 1$)	30 ($c = 2$)	MRI ($c = 1$)	
R3	Shape ($c = 2$)	Corpus Callosum ($c = 3$)	Schizophrenia ($c = 1$)	50 ($c = 1$)	MRI ($c = 1$)	
Importance Coefficient Set 1	1	0.5	0.5	0.2	0	∑ R1 = 3.6 ∑ R2 = 3.9 ∑ R3 = 4.2
Importance Coefficient Set 2	0.5	0.5	0.2	1	0	∑ R3 = 3.7 ∑ R2 = 4.2 ∑ R1 = 4.6

Table 1. Finding similar studies. For each related feature (first line), the three records listed (R1, R2, R3) contain a value and corresponding result from the choice function c (lower values indicate higher ranking). Two importance coefficient sets are given to show two different overall rankings (in the bottom right cells; again, lower values indicate higher overall ranking).

number of items), it is possible to customize a search ranking for the particular aspects a user is looking for. One might put “hard” limits on certain meta-data (for example requiring fairly recent entries) and put preferences on others (for example expressing a strong or weak preference for the number of items being large). The customized ranking r is formulated by assigning to each data item x a linear combination of meta-data “choice functions" c_i with associated *importance coefficients* α_i, where α_1, …, α_k are specified by the user and α_0 is a coefficient provided by the system to incorporate feedback from previous searches:

$$r(x) = \sum_{i=0\ldots k} \alpha_i c_i(x) \qquad (1)$$

Choice functions are functions of meta-data that select a range on the particular ordering appropriate for each meta-data type; for example, the most common choice function for date is “newer than …", and the most common for size (number of subjects) is “of size at least …". The importance coefficients attached to each are discretized to a small number of values, each of which can be represented by a word or phrase for user input: “necessary” (1), “very important” (0.8), “important” (0.5), “not very important” (0.2), and “doesn't matter” (0).

For example, a user wants to find similar studies about ‘*volumetric analysis of hippocampus structure in schizophrenia using MRI scans and involving as many subjects as possible*’ (Table 1). If the user is more interested in the “volumetric analysis”, he/she can apply scheme 1 and system will return R1 as the best matched result. On the other hand, if he/she is more interested in the size of study set, scheme 2 can be applied and the system will return R3 as the best matched result.

III. A MetaDL EVALUATION METRIC

Evaluation of DLs [4] can be considered in the system (e.g., interoperability, scalability, heterogeneity, reliability, and integration), user (e.g., relevance, specificity, timeliness, effort vs. effect, and usability), and content (e.g., sufficiency, currency, and quality) domains. The evaluation addressed in this paper resides in the user and content domains; specifically, how closely MetaDL contents match the interests of users. We propose a measure of this match, based on the structure of advanced meta-information queries like those described for the BrassDL system in Section II.

Because a MetaDL by definition includes standardized meta-data for each item, searches involving these meta-data are easier to compare than in the general case. "Basic" searches, e.g. keyword searches, are not considered here. For advanced searches, prioritizations can be aggregated and considered independently of particular searches, resulting in an overall measure of which meta-data (and thus which data) are most relevant to most users. In BrassDL, for example, we might find that for "data type" fMRI scans are preferred to PET, or that age of data is (or isn't) an important factor for users.

The evaluation of data relevance as described is only an approximation, since it generalizes queries done with combinations of meta-data to each meta-data constituent independently of the others. A more refined variation of this is to consider sets of k meta-data items, $k \leq n$, where n is the total number in the system, in combination. For $k = 1$, this is identical to what was described above. For $k = 2$, it considers the relative importance users attach to *pairs* of meta-data items. Extending the example above, this might allow us to determine, for example, that PET data is very popular for studies of schizophrenia, and that old data is more often sought for cognitive tasks. The framework also provides a numerical ranking of each of these, which allows them to be compared to each other.

It is not reasonable to calculate every combination for every value of k, since the number of combinations increases factorially: $N_k = \frac{N!}{(N-k)!}$. However, it is still possible to consider and compare all combinations below a threshold value (e.g., $k = 3$), and it may be informative to determine the highest-valued combination for some higher values of k.

The formulas below calculate an R value that evaluates *how important* a metadata item (e.g. datatype) is on average. Let N be the number of advanced search queries, N_k the size of the search group that includes search k, G the number of search groups (an obvious use of this is one group per user, but can be any grouping), $\alpha_{i,j}$ the prioritization of metadata item i in search j.

Evaluation of a metadata item: $R(i) = \frac{1}{N}\sum_{s=1}^{N}\alpha_{i,s}$

Same evaluation with weighting for number of user queries, if available: $R^*(i) = \frac{1}{G}\sum_{s=1}^{N}\frac{\alpha_{i,s}}{N_s}$

Evaluation of a pair of metadata items: $R(i,j) = R(i)R(j)$

Evaluation of any number, k, of metadata items: $R(1,\ldots,k) = \prod_{i=1}^{k} R(i)$

Prioritizations of meta-data parameters in an advanced search can take on several different values, and so in the latter case is necessary to add a second parameter to the evaluation of data relevance that we calculate: a mapping for the importance coefficients allowed in expert search (Section III). With this second parameter, it is possible to formulate evaluations of popularity like the following: “what is the most common combination of meta-data requirements of 'very important' or higher” or “what is the most often required meta-data value”. It is also possible to add as additional parameters cutoff values for the expert search choice functions, resulting in the ability to evaluate “what is the distribution of user requirements for recency for different data”, or “what is the most preferred minimum size for PET datasets”.

The choice function cutoffs are necessarily different for every metadata item, since they all have different ranges (and data types: integer, floating point, or text). Because of this, we use a set θ of fixed cutoff parameters (one per metadata item) and a function f_θ that filters according to cutoff:

$$f_\theta(\alpha) = \begin{cases} \alpha & \text{if } \alpha > \theta \\ 0 & \text{otherwise} \end{cases}$$

Then the evaluation formula, *R*, is as before except for the application of the filter f_θ:

$$R(i) = \frac{1}{N_\alpha} \sum_{s=1}^{N} f_\theta(\alpha_{i,s})$$

and likewise for R*, where N_α is the number of queries with α over the threshold θ. This allows one to evaluate, e.g., “what are the most often required ($\alpha = 1$) metadata items?” The result is a numeric ranking of metadata items, and as described previously, for each a histogram can be calculated giving the proportion of each value that was required (e.g., “PET” or “fMRI” for datatype).

IV. CONCLUSIONS

We have presented BrassDL, a MetaDL for human brain data that incorporates highly heterogeneous imaging data and results, and an evaluation metric that allows the user to easily narrow a search to those data sets that best match the user’s needs.

MetaDLs are timely and important in the field of neuroscience to support research-activity documentation, scientific discovery, collaboration and sharing of extremely valuable data. As new imaging technologies develop, the field is exploding with new, highly valuable data that amass in individual laboratories with public funding and which are basically non-shareable. Access and combination of these data will lead to very large sets that connect the function and structure of human brain, furthering the goal of the Human Brain Mapping Project.

Evaluation of expert search popularity is a useful tool in evaluating how a MetaDL corresponds with its users needs.

REFERENCES

[1] F. Makedon, J. Ford, L. Shen, T. Steinberg, A. Saykin, Heather Wishart, and S. Kapidakis, “MetaDL: A Digital Library of Metadata for Sensitive or Complex Research Data”, submitted to ECDL, Sept. 2002, Rome, Italy.

[2] Fox, P. T. and Lancaster, J. L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*, 3(4):319-321.

[3] Grethe, J. S., Van Horn, J. D., Woodward, J. B., Inati, S., Kostelec, P. J., Aslam, J. A., Rockmore, D., Rus, D., and Gazzaniga, M. S. (2001). The fMRI data center: An introduction. *NeuroImage*, 13(6):S135.

[4] Ronald L. Larsen. The DLib Test Suite and Metrics Working Group: Harvesting the Experience from the Digital Library Initiative. *D-Lib Working Group on Digital Library Metrics Website*, April 2002.
http://www.dlib.org/metrics/public/papers/The_Dlib_Test_Suite_and_Metrics.pdf

[5] METAe Metadata Engine. http://meta-e.uibk.ac.at/

[6] MUMIS: Multimedia Indexing and Searching Environment, University of Twente. http://parlevink.cs.utwente.nl/projects/mumis.html

[7] Dublin Core Metadata Initiative. http://www.dublincore.org/

[8] C. Houstis, S. Lalis, "ARION: An Advanced Lightweight Software System Architecture for accessing Scientific Collections", Cultivate Interactive, issue 4, May 2001

[9] J. Ford, H. Farid, F. Makedon, L. A. Flashman, and A. J. Saykin, "Classification of fMRI Brain Activation Maps", in preparation.