

DOPAMINE – A tool for visualizing clinical properties of generic drugs

C.J. Wroe¹, W.D. Solomon¹, A.L. Rector¹ and J.E. Rogers¹

Abstract. Visualization tools are becoming more important as ontologies increase in size and complexity. One example of a large ontology is the Drug Ontology developed at the Medical Informatics Group using GALEN classification techniques. It will provide a reference terminology for PRODIGY; a project providing computerized medicine prescribing guidelines for primary healthcare in the UK. While developing the Drug Ontology we have needed a customized visualization tool for authoring and checking ontology entries. The software tool (*Drug Ontology Production And Maintenance Environment or DOPAMINE*) provides a tabular summary of sets of ontology entries.

DOPAMINE has improved the consistency and speed of authoring entries, and provided a means of highlighting, previously hard to detect, authoring errors. Although specific to the drug ontology, the techniques used could be generalized for other ontologies.

1 INTRODUCTION

Tabular summaries are an effective means of visualizing structured data, and are the mainstay of database and spreadsheet user interfaces. Several research groups have used tables to present summaries of knowledge base entries.

The Generic Knowledge Base Editor (GKB – Editor) is being developed at SRI International's Artificial Intelligence Center. It is a tool for 'graphically browsing and editing knowledge bases across multiple Frame Representation Systems (FRSs) in a uniform manner' (see [1]). It has a spreadsheet viewer facility but is limited to instances of a knowledge base. The Drug Ontology in contrast describes only classes of drug concepts, such as 'All drugs containing atenolol hydrochloride'.

Conceptually Oriented Description Environment (CODE4) has been developed at the Department of Computer Science, University of Ottawa [2]. Using a 'Property Comparison Matrix', it is possible to visualize the similarities and differences between the properties of concepts. These functions are very similar to those needed by DOPAMINE. A related display is the 'Property Inheritance Matrix' which displays values of a properties as they change down a subsumption hierarchy. The tool can be used for both retrieval and authoring of information.

Protégé is an environment developed at Stanford Medical Informatics for building reusable ontologies and problem solving

methods [3]. The environment includes the ability to use a table widget. The table widget is designed to simplify authoring of instances in a knowledge base [4]. The table is regarded as an integral part of the knowledge base, rather than an independent method of visualising the knowledge.

Within our department, Gary Ngg, a PhD student, has been developing tools for visualizing ontologies authored in Grail. Tables show summaries of Grail properties. 'Lenses' can then be superimposed over the table to visualise the result of user defined queries on the table [5].

The development of DOPAMINE was born out of a necessity to provide a user interface with which large and verbose drug ontology class descriptions could be authored and checked easily and rapidly. It draws on previous research on tabular based summaries and applies it to the specific task within the Drug Ontology project.

1.1 Rational behind the drug ontology

The Prodigy Project is developing prescribing guidelines for UK primary healthcare [6]. These guidelines are triggered when the doctor enters a diagnosis, and guide the doctor through prescribing decisions. The project is now in its third phase and is developing more complex guidelines for chronic diseases. These guidelines are active over several months of a patient's chronic condition and need to reference terms in the patient record, for example, to detect the patient's current medication. With this in mind, they recognized the need for a scalable terminology solution.

Through the GALEN and GALEN In USE programs [7], we have been developing methods for creating and maintaining scalable terminologies in the medical domain. The work centres on the use of ontologies - explicit formal representations of concepts [8]. These allow communication, reuse, and representation of medical concepts in a logical system [9]. A description logic, named Grail, has been used to implement the ontology [10]. Description logic provides classification and consistency services for the ontology. To date, ontologies have been created that cover anatomy, basic physiology, basic pathology and basic medical devices. Together these form the Common Reference Model (CRM) [11].

The 'Drug Ontology' builds on existing ontologies of pathology and physiology to create formal descriptions of a generic drug's clinical properties. These properties comprise: ingredients, formulation, indications, contraindications, cautions, mechanism of action, interactions, side effects and clinically relevant pharmacokinetics [12].

¹ Medical Informatics Group, Department of Computer Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK, email: cwroe@cs.man.ac.uk & Open Galen Organisation, <http://www.OpenGalen.org>.

The project is producing descriptions of primary care relevant generic drugs in the British National Formulary (BNF) [13]. There are approximately 1500 such entries in the BNF. During the project we have found that the explicit formal ontology descriptions, soon become verbose and difficult to read by human readers. Information has to be included in the descriptions that human readers would automatically infer, and so appear redundant.

1.2 Intermediate Representation

To help simplify authoring ontology descriptions an Intermediate Representation (IR) was previously developed in the GALEN in USE Project. IR is a simplified formal language used to describe the definition and properties of a concept [14][15]. The IR is automatically translated into the lower level language Grail, with which the description logic classifier operates. Translation rules are authored for each domain, allowing some degree of ambiguity in each Intermediate Representation.

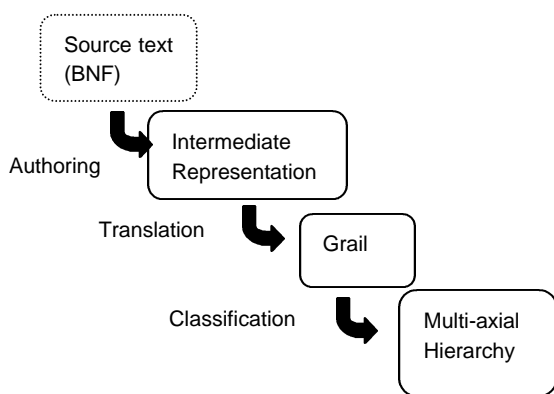


Figure 1. Processes involved in producing a drug ontology description using intermediate representation.

Each term description (called a dissection) starts with the keyword MAIN which is interpreted in a domain specific way. For example, in the drug ontology, it is translated to Grail as ‘Drug which hasIngredient’.

A set of terms (called descriptors) and semantic links, follow the MAIN keyword, which specify the terminological definition and properties of the concept being described. Indentation of the links is used to specify which descriptors are being linked.

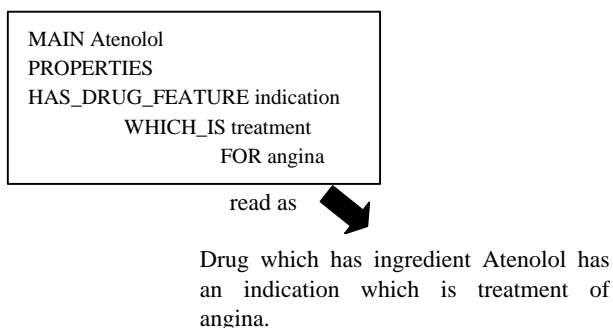


Figure 2. How Intermediate Representation should be interpreted.

If ambiguity exists *within* a domain, as with the term ‘indication’ in the drug domain, the exact meaning of the term has to be specified at the level of the Intermediate Representation. As the number of properties of a drug increases it can become difficult to read (see Figure 3).

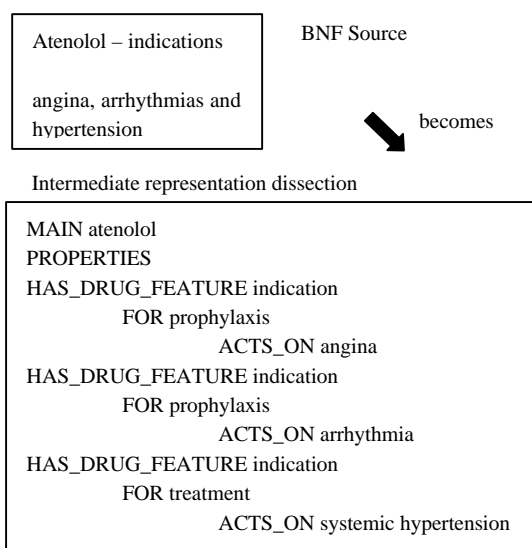


Figure 3. The result of authoring an Intermediate Representation dissection for a sample of source text.

2 METHOD

2.1 Define queries

The description of each type of property follows a stereotyped pattern. For example, all indication properties begin ‘HAS_DRUG_FEATURE indication FOR ..’. The first step is to manually author a query. This will be used by the application to organise properties of the same type, for example, all indications. These queries are specified using a purpose built entry tool.

2.2 Execute queries

The user selects a group of drug descriptions to be analyzed. For example, the set of all beta-adrenoceptor blocking drugs. The drug descriptions within the set are checked to see if any properties match one of the predefined queries. If a property does match it is added to a list of detected properties to be presented to the user. For example an indication query will detect and collate all properties beginning with the structure ‘HAS_DRUG_FEATURE indication FOR ..’

2.3 Present results as a view with which the user can interact

A table is constructed with a list of properties detected along the vertical axis and the list of drugs examined on the horizontal axis. The list of properties is grouped by the queries that the user has previously defined. Properties are further grouped by common initial structure. This process is explained in figure 4. The top table shows the view that would be presented if the common structure was repeated. The bottom table is an extract from the DOPAMINE tool and shows indications for Atenolol. Users can also add to drug ontology descriptions using the tool.

Properties	Atenolol
Indication	
FOR prophylaxis	
ACTS_ON angina	X
Indication	
FOR prophylaxis	
ACTS_ON arrhythmia	X
Indication	
FOR treatment	
ACTS_ON hypertension	X

Condensing common structure

2468 ATENOLOL	
ATENOLOL	
Features:	
- indication	3
FOR adjunct	
FOR induction	
FOR maintenance	
FOR management	
FOR premedication	
- FOR prophylaxis	2
ACTS_ON arrhythmia	1
ACTS_ON angina	1
FOR tranquilisation	
- FOR treatment	1
ACTS_ON hypertension	1

Figure 4. Presentation of properties in a condensed format.

2468 ATENOLOL	
ATENOLOL	
Features:	
- indication	3
FOR adjunct	
FOR induction	
FOR maintenance	
FOR management	
FOR premedication	
- FOR prophylaxis	2
ACTS_ON arrhythmia	1
ACTS_ON angina	1
FOR tranquilisation	
- FOR treatment	1
ACTS_ON hypertension	1
HAS_FEATURE licensed	1

Represents

HAS_DRUG_FEATURE indication

 FOR prophylaxis

 ACTS_ON arrhythmia

HAS_DRUG_FEATURE indication

 FOR prophylaxis

 ACTS_ON angina

HAS_DRUG_FEATURE indication

 FOR treatment

 ACTS_ON hypertension

HAS_FEATURE licensed

Figure 5. Extract from a DOPAMINE view showing indications of Atenolol with the use of bars.

This method of presentation produces a problem if the structure of the property includes two links attached to the same descriptor. An additional notification is necessary to distinguish multiple

semantic links attached to one descriptor in contrast with two separate properties that have a common initial section. This is achieved by the use of bars. These group together semantic links that belong to the same property (see figure 5 for an example). When authoring with the tool it is necessary for the user to indicate whether a bar is necessary.

3 EVALUATION

3.1 Authoring

The Drug Ontology project is producing drug descriptions of primary care relevant generic drugs in the BNF. Initial authoring of the 1500 descriptions is now complete. The 'DOPAMINE' authoring tool was developed after initial authoring experience and was operational mid-way through the project. It has therefore been used to author 700 descriptions. The average size of a description is 21 properties.

The major rate-limiting step in authoring with the tool was creating the left-hand candidate property list. A property can only be added to a description if it appears along the vertical axis of the table. The source text is semi-structured containing a large number of comma-separated lists. A major increase in authoring speed was therefore gained by simply separating candidate terms using punctuation.

BNF source text descriptions of drugs in the same therapeutic class, often share sets of properties such as contraindications. A mechanism therefore had to be added which allowed copying of sets from one description to another. The visual representation of the information provided users with instant confirmation that the process had produced the correct result.

3.2 Checking

Of the 1500 descriptions, 300 have been released to the PRODIGY guideline developers for external appraisal. Before release these descriptions were checked using the DOPAMINE tool. Therapeutic classes of drugs were examined as individual sets. Their properties were compared and checked against the source text.

The tool allowed visualization of information not readily apparent in the textual representations of the descriptions.

Previously, visualizing the classification of drug concepts was the main mechanism for checking the validity of drug descriptions. Users found it easy to detect mis-classifications based on incorrect ontology descriptions. For example, 'Atenolol tablet' clearly should not be a child of 'Atenolol injection'. However, it was much more difficult to detect *missed* classification due to omissions in ontology descriptions. For example, 'Atenolol intravenous injection' should be a child of 'Atenolol injection' but may in fact have been placed in another area of the hierarchy due to an omission in its description.

Using the tabular view, the authors were able to detect omission errors, not detected by checking the classification of drug concepts. These errors showed up as sparse areas in the table for one drug compared with a dense corresponding area for a neighbouring related drug. The omission errors were often due to a failure to copy a set of properties from one drug description to another.

Using the tabular view, the authors were also able to detect multiple use of similar terms in the source text to represent the same property.

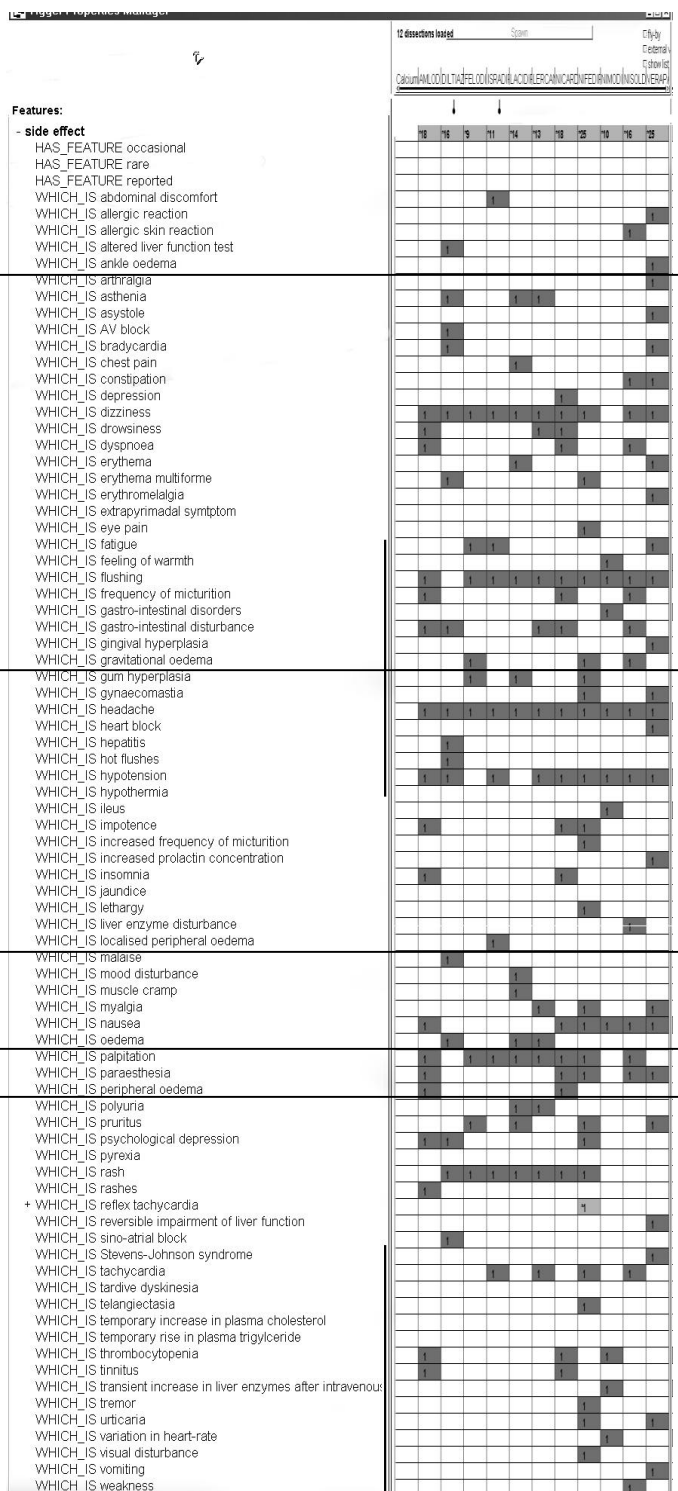


Figure 6. Extract of a DOPAMAINE view showing side effects of calcium channel blockers. Columns show properties of the calcium channel blocker class description, Amlodipine, Diltiazem, Felodipine, Isradipine, Lacidipine, Lercanidipine, Nicardipine, Nifedipine, Nimodipine, Nisoldipine, Verapamil. Horizontal lines have been added to highlight similar terms.

In the calcium channel antagonist section, five related terms have been used in the source text to state that specific calcium channel blockers have a ‘peripheral oedema’ side effect (‘Ankle

oedema’, ‘gravitational oedema’, ‘localized peripheral oedema’, ‘peripheral oedema’, and ‘oedema’) (see Figure 6).

Additional organization to group similar terms together is discussed later (see section 4.1), but even with this alphabetical organization of concepts it is possible to identify this possible inconsistency in the drug descriptions. Although inconsistencies of this nature may not be important for professional readers of the source text, automated decision support relies on consistency of terms for correct operation.

As a complete therapeutic class of drugs can be visualized, the table can be used to detect properties which are true for all members. In this case it may be useful to promote this property to the class description and reduce redundancy [16].

Figure 7 shows that all members of the clofibrate class possess breast feeding and pregnancy contraindications. These contraindications could therefore be promoted to the description of the clofibrate drug class.

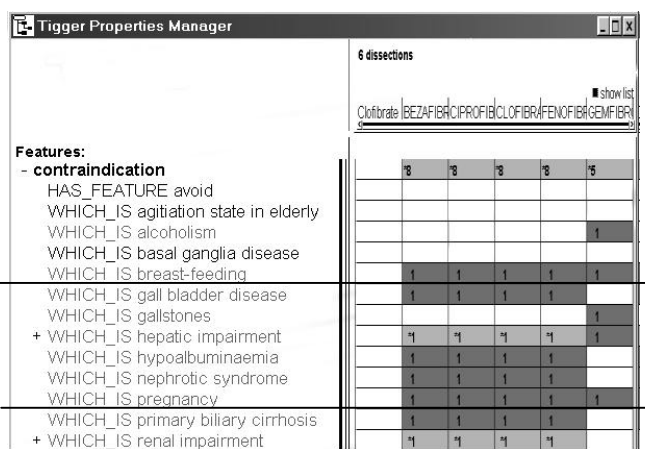


Figure 7. Extract of a DOPAMAINE view showing contraindications of the clofibrate class of drugs. Columns show properties of clofibrate class, Bezafibrate, Ciprofibrate, Clofibrate, Fenofibrate and Gemfibrozil. Horizontal lines have been added to highlight properties true of all class members.

4 DISCUSSION

4.1 General tool design

At present, properties are organized based solely on the information in the structure of the IR drug descriptions. If a drug has a list of 30 side effects that do not differ in structure but only in the terms used, they will only be organized as a flat list (see figure 6). In the initial stages of authoring this is the only information available. However, as the terms in the drug descriptions are mapped to concepts in the Common Reference Model, a more detailed conceptual organisation is possible. Conceptually similar terms will be grouped under common parents. For example the flat list of 30 side effects could be grouped by which body system they involve. Further work is needed to make this organisation possible and explore any difficulties of visualisation raised with a multi-axial organisation.

The structure of the queries used to group related features is very specific to this application. To allow use in other domains the specification of queries would need to be made more general. This may affect the way properties are displayed and their impact would need to be explored.

4.2 Authoring

Extracting terms from the source text and placing them on the vertical axes of the table was the major factor which determined speed of authoring. While it is not the aim of this project to completely automate the extraction process, the semi-structured nature of the source text could be used to more effectively provide a small candidate list of terms from which the author can choose.

Visualization of properties is most effective when working on one therapeutic class of drugs. Selection of the set is entirely manual at present, which could lead to a drug description being incorrectly missed out of a set. An automatic query or classification-based selection may be more reliable.

4.3 Checking

Errors can arise when the IR descriptions are translated into Grail and classified. As the focus of the project moves away from bulk authoring to bulk validation, it will be necessary for the tool to use the compiled Grail model as a source for visualization. Tools have already been developed to reverse the translation process and produce IR from Grail. It is therefore possible to provide a view based directly on the underlying Grail representation. Future work involves testing this approach.

5 SUMMARY

For automatic classification of drug terms to be successful the formal description of those terms needs to be unambiguous and so verbose. This limits the productivity of authors. We have produced more concise views for authors to interact with and in doing so have provided novel opportunities to visualize clinical drug information. These views have been successfully used to both author the drug descriptions and increase the consistency of the information within those drug descriptions. As the amount of information about each drug grows, even these views are becoming too complex. Further work is needed to make use of the domain information that is available in the related ontologies of the Common Reference Model to further organise the presentation of drug properties.

ACKNOWLEDGEMENTS

The Drug Ontology Project is funded by the NHS Executive. We would like to thank the PRODIGY team at the Sowerby Center For Health Informatics, University of Newcastle and the British National Formulary for their support.

REFERENCES

- [1] GKB-Editor Home Page.
URL: <http://www.ai.sri.com/~gkb/welcome.shtml>
- [2] Doug Skuce, Timothy C. Lethbridge. *CODE4: a unified system for managing conceptual knowledge*. International Journal of Human-Computer Studies, Vol. 42, No. 4, Apr 1995, pp. 413-451
- [3] W. E. Grosso, H. Eriksson, R. W. Fergerson, J. H. Gennari, S. W. Tu, & M. A. Musen. *Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000)*. 1999. URL: http://smi.web.stanford.edu/pubs/SMI_Abstracts/SMI-1999-0801.html
- [4] Protégé 2000 Tutorial.
URL: <http://smi-web.stanford.edu/projects/protege/protege-2000/doc/tutorial/index.html>

- [5] Gary Ngg. *Interactive Visualisation Techniques For Ontology Development*. PhD Thesis submitted for examination at the University of Manchester UK, May 2000.
- [6] I Purves. *Prodigy interim report*. Journal of Informatics in Primary Care 1996:2-8.
- [7] A.L. Rector, P.E. Zanstra, W.D. Solomon, J. E. Rogers et al. (1998). *Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine*. IEEE Transactions on Information Technology in BioMedicine, Special issue on the EU Healthcare telematics programme; Vol. 2 No.4 pp. 229-242
- [8] N. Guarino., P. Giaretta. *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. In N. J. I. Mars (ed.), *Towards Very Large Knowledge Bases*, IOS Press 1995.
- [9] Mike Uschold, Martin King. *Towards a Methodology for Building Ontologies*. Technical Report AIAI-TR-183, Artificial Intelligence Applications Institute, Edinburgh University, UK, 1995.
- [10] A.L. Rector, S.K. Bechhofer, C.A. Goble, I. Horrocks, W.A. Nowlan, W.D. Solomon. *The GRAIL Concept Modelling Language for Medical Terminology*. Artificial Intelligence in Medicine, Volume 9, 1997.
- [11] Open Galen Common Reference Model
URL: <http://www.openGalen.org>
- [12] W.D Solomon, C.J. Wroe, J.E Rogers, and A.L. Rector. *A reference terminology for drugs*. Journal of the American Medical Informatics Association 1999(1999 Special Conference Issue).
- [13] British National Formulary. British Medical Association and the Royal Pharmaceutical Society of Great Britain, 2000.
- [14] J.E. Rogers, W.D. Solomon, A.L. Rector, P.E. Zanstra. *From rubrics to dissections to GRAIL to classifications*. Medical Informatics Europe (MIE-97). Thessalonika, Greece: IOS Press, 1997:241-245.
- [15] W.D. Solomon, J. Rogers, A. Rector, E J van der Haring, P.E. Zanstra. *Supporting the use of the GALEN Intermediate Representation*. Journal of the American Medical Informatics Association 1998(1998 Special Conference Issue).
- [16] C.J. Wroe, W.D. Solomon, J.E. Rogers, A.L. Rector. *Inheritance of Drug Information*. Submitted to Journal of the American Medical Informatics Association 2000 (2000 Special Conference Issue).