DOPAMINE – A tool for visualising clinical properties of generic drugs

C.J. Wroe[1],W.D. Solomon[1], A.L. Rector[1], J.E. Rogers[1]
[1]Medical Informatics Group,
Department of Computer Science, University of Manchester, UK
& OpenGALEN www.opengalen.org

The Drug Ontology Project has been developed to provide a reference terminology by which prescribing guidelines within the PRODIGY Project and the patient record can be integrated. Consistency and ease of maintenance are provided by GALEN classification techniques. These classification techniques rely on full and unambiguous formal descriptions of drug terms. We have found during the project that these descriptions become verbose and difficult to author and maintain in their native format. To manage the problem we have developed the '*D*rug *O*ntology *P*roduction *A*nd *Ma**I*ntenance e*N*vironm*E*nt' or DOPAMINE. DOPAMINE presents concise views of a drug's clinical properties with which the user can interact. We have found these tools not only simplify authoring but also provide new opportunities for visualising information that is not possible with the formal descriptions alone or the original source text.

Introduction

The Drug Ontology is aiming to provide a reference terminology of drugs to support prescribing guidelines developed within phase III of the PRODIGY Project in the UK[1,2]. The first two phases focused on prescribing for acute conditions within a single consultation. Phase III is concentrating on providing prescribing support for chronic diseases. As the guidance will have to apply over multiple consultations the guideline system must interact much more with the patient record.

Guidelines are authored using terms that need to apply to a population of individuals and so are general such as 'anti-anginal drug'. In contrast, the patient record contains information about actual prescribing events, and so includes very specific drugs terms such as 'Atenolol 25mg, bd, prescribed 21/7/2000'. A mechanism is needed to relate the two sets of terms if the guideline system is to interact with the patient record.

Traditionally, classifications have provided the link. However, existing classifications have been manually created to fulfil specific tasks. For example the classification of the British National Formulary (BNF)[3] has been created primarily for manual navigation from abstract drug classes to specific drug monographs. Decision support is a novel task, and guideline authors have found existing classifications do not provide the necessary abstract drug classes such as 'anti-anginal drug'. As they are manually created, it is a labour-intensive task to create and maintain the relationships of these new abstractions.

GALEN classification techniques have been developed in other medical domains such as surgery[4]. Classification is achieved automatically using a description logic classifier. Automatic classification depends on a formal description of each term with which the classifier can infer the relationships between the terms. The effort is therefore shifted from manual classification to explicit definition of the terms present in the classification. The Drug Ontology project involves writing formal descriptions for generic drug terms such as 'Atenolol' together with more abstract terms such as 'Anti-anginal medication'. These are then presented to the classifier to produce a classification of drug terms. We have shown the advantages to this approach are: [5,6,7]

1) Information by which the classification has occurred is explicit.
2) The classification is logically sound.
3) The exact organisation of the classification can be tailored to different tasks by using the same generic drug term definitions with differing abstract terms.

Each description must be as unambiguous as possible. The exact meaning of semantic links such as 'indication' need to be expanded if the classification is going to be used independently of the source text. In this case 'indication' has to be expanded to 'indication as suggested by BNF' and further specified by the goal of treatment such as 'curing', 'palliating', 'preventing'.

Descriptions must also contain sufficient information to allow classification along the axes required by the guideline authoring team. For example calcium channel blockers may need to be classified by chemical structure, mechanism of action and indications. The calcium channel blocker description would therefore have to describe all these properties. Although properties such as indications are related to mechanism of action each must be specified separately to allow classification independently along each axis.

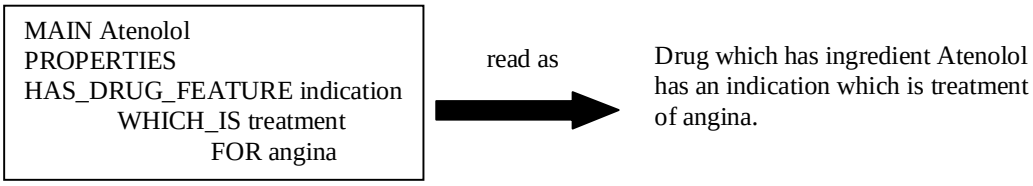
Unambiguous descriptions become verbose

During the project we have found that unambiguous descriptions, soon become verbose in the eyes of human readers. Information has to be included in the drug term descriptions that human readers would automatically infer, and so appear redundant.

An introduction to Intermediate representation, the formal language used to write drug descriptions.

The GALEN Intermediate Representation is the simplified formal language used to describe the definition and properties of a concept[8]. Each term description (called a dissection) starts with the keyword MAIN which is interpreted in a domain specific way. 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 1. How Intermediate Representation should be interpreted.



The intermediate representation is then automatically translated into a lower level description logic language GRAIL with which the classifier operates[9]. Translation rules are authored for each domain allowing some degree of ambiguity in each Intermediate Representation. However if ambiguity exists *within* a domain such 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.

Figure 2. Processes involved in producing the drug ontology.

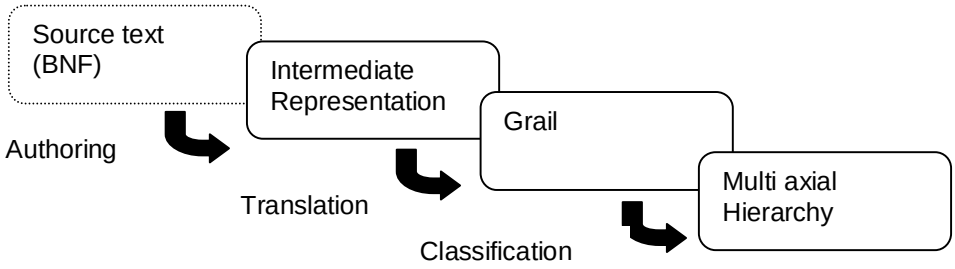


Figure 3. The result of authoring an Intermediate Representation dissection for a sample of source text.

BNF Source text

Atenolol – indications

angina, arrhythmias and hypertension

becomes

Intermediate representation dissection

```
MAIN atenolol
PROPERTIES
HAS_DRUG_FEATURE indication
        FOR prophylaxis
                ACTS_ON angina
HAS_DRUG_FEATURE indication
        FOR prophylaxis
                ACTS_ON arrhythmia
HAS_DRUG_FEATURE indication
        FOR treatment
                ACTS_ON systemic hypertension
```

Although the unambiguous form is needed for correct classification, authors need to interact with a more concise view of the descriptions. 'DOPAMINE' is the name of the toolset developed for presenting concise views of drug descriptions to users.

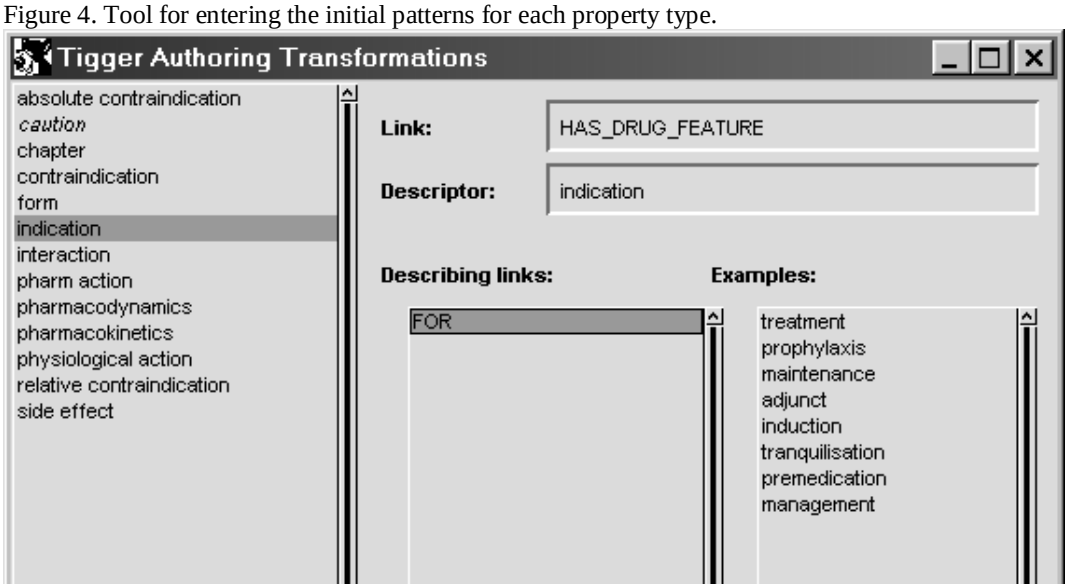
Method

1) Define repetitive patterns.
There are nine main clinical drug properties the 'Drug ontology' describes.

- Ingredients
- Form
- Indications
- Cautions
- Contraindications
- Effects
- Side effects
- Interactions
- Pharmacokinetics

The description of each type of property follows a stereotyped pattern. This pattern is specified manually for each type of property. Using this pattern, DOPAMINE is able to recognize the repetitive initial pattern of properties and so provide a more concise view to the user. The following diagram shows the pattern defined for indication properties.

Figure 4. Tool for entering the initial patterns for each property type.



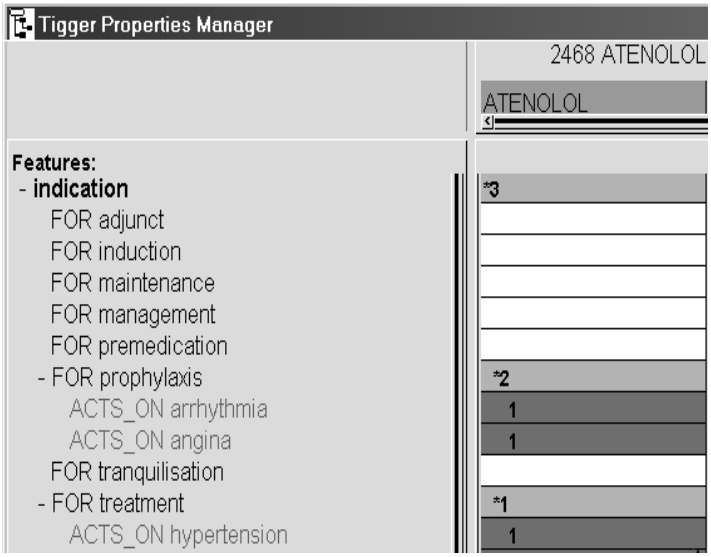
2) Parse verbose description for occurrences of these patterns.
Each property of a drug description is checked to see if it matches one of the predefined patterns. Usually a set of drug descriptions is presented together under user control. For example the set of all beta-adrenoceptor blocking drugs.

3) Present results as a view with which the user can interact.
A table is constructed with a list of properties detected along the Y-axis and the list of drugs examined on the X-axis. The list of properties is grouped by the patterns that the user has previously defined. If two properties have a common initial section, only the variations are listed below the common structure. This process of is explained in figure 5. The table on the right is the view that would be presented if the common structure was repeated. The section of the table on the right is from the DOPAMINE tool and shows indications for Atenolol.

Figure 5. Presentation of properties in a condensed format.

Properties	Atenolol
indication	
FOR prophylaxis	
ACTS_ON angina	X
indication	
FOR prophylaxis	
ACTS_ON angina	X
indication	
FOR treatment	
ACTS_ON hypertension	X

Condensing common structure



The numbers in the boxes indicate how many properties include that initial section. An asterisk indicates that this structure is only a preliminary part of those properties. For example, Atenolol possesses three properties that begin 'HAS_DRUG_FEATURE indication'. Two of these continue with 'FOR prophylaxis' and one continues with 'FOR treatment'. The three properties end with 'ACTS_ON arrhythmia, 'ACTS_ON angina, and 'ACTS_ON hypertension' respectively.

What happens if the structure contains two terms at the same indentation level?
This method of presentation produces a problem if the structure of the property includes two links attached to the same descriptor term. An additional notification is necessary in the view to distinguish multiple links from two separate properties that have a common initial section.

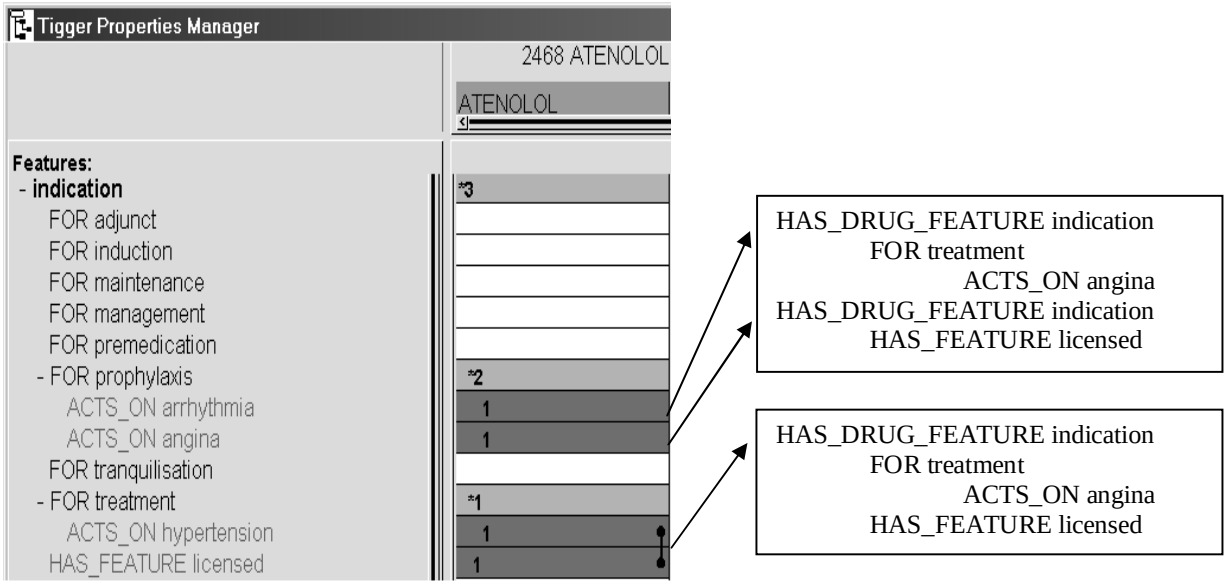


Figure 6. Extract from DOPAMINE display showing indications of Atenolol with the use of bars.

The bar between properties is used to distinguish one property possessing two links at that level, as opposed to two different properties condensed into one view.

Structured descriptions provide opportunities for visualizing patterns not possible in the source text.

Given that the clinical drug information is now in a highly structured form and a concise view can be presented to the user, the tools provide new opportunities for visualizing patterns of information. The tool can present a group of related drugs such as betablockers in one table. The user can then identify significant patterns of properties that need further attention.

The view has been used to detect if a group of properties is missing from one drug description by recognizing a blank area in the table. It has also been used to determine whether a property is true for all members of a drug class and so move that property to the description of the drug class.

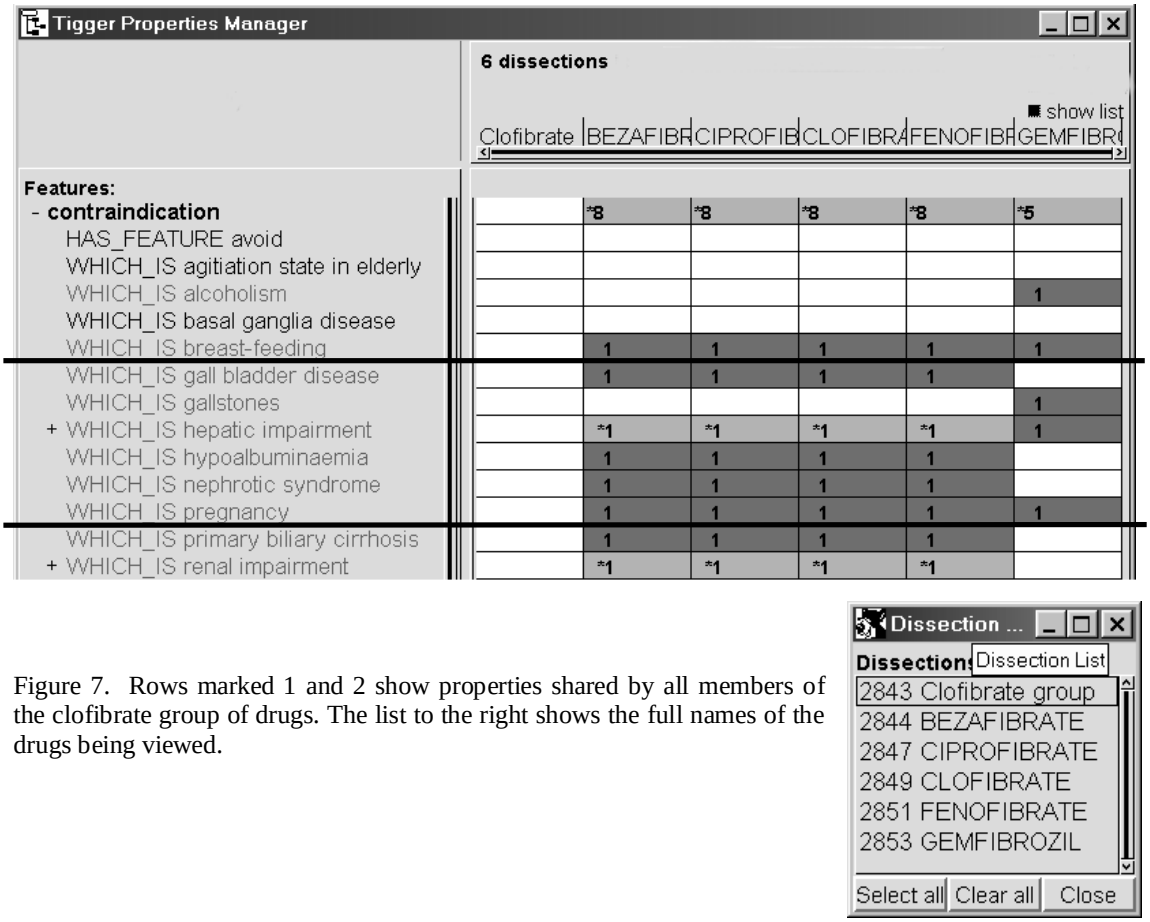
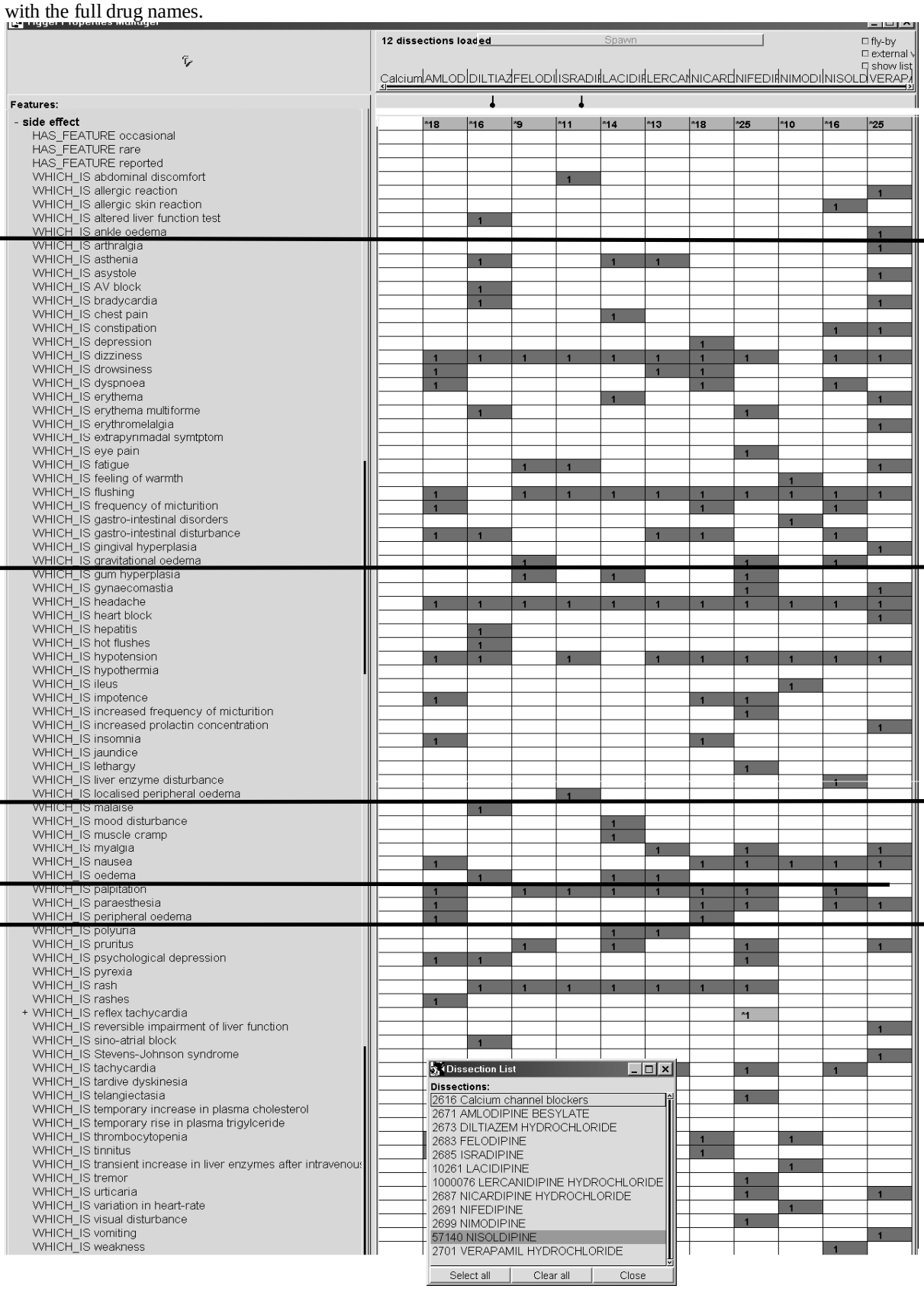


Figure 7. Rows marked 1 and 2 show properties shared by all members of the clofibrate group of drugs. The list to the right shows the full names of the drugs being viewed.

In the above example 'breast feeding' and 'pregnancy' contraindications are present for all members of the 'Clofibrate group' of drugs and so it may be more appropriate to attach this property to the description of the 'Clofibrate group drug class'. Once classification had occurred these properties would then be inherited by the members of the class[10].

The figure 8 shows how it has been used to detect whether different terms have been used to describe the same property in a set of related drugs. Five related terms have been used in the source text to describe the 'peripheral oedema' side effect possessed by calcium channel blockers: 'Ankle oedema', 'gravitational oedema', 'localized peripheral oedema', 'peripheral oedema', and 'oedema'. Additional organization to group similar terms together is discussed later, 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.

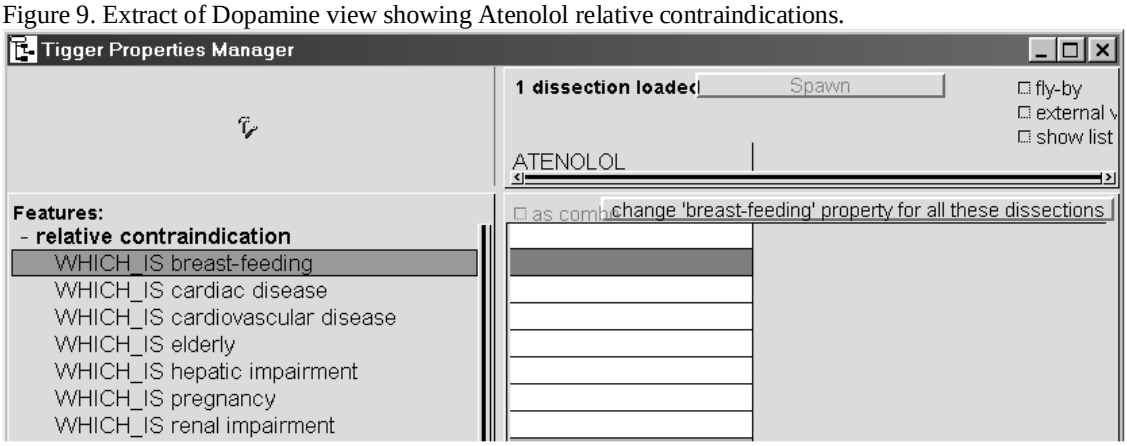
Figure 8.. Extract of DOPAMAINE view showing side effects of calcium channel blockers. A list is present with the full drug names.



The tool can also be used for authoring with associated tools.

The table is interactive allowing authoring of descriptions using the toolset.

Figure 9. Extract of Dopamine view showing Atenolol relative contraindications.



Pressing the 'change 'breast feeding' property for this dissection' will add the relative contraindication for Atenolol resulting in the addition of the following intermediate representation:

'HAS_DRUG_FEATURE relative contraindication
WHICH_IS breast-feeding'

Selecting more than one drug allows rapid authoring of consistent information for a group of descriptions.

Discussions and further work

At present the way the properties are organized is based solely on the information in the structure of the 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 in a flat list. Figure 8 shows that users need more organization of these lists if they are to easily find the relevant property they are looking for. The pattern of drug properties will also become more meaningful if conceptually similar terms are grouped. The terminological information to produce this organization is available because is it also needed for the automatic classification process. For example the classifier needs to recognize that 'angina' is a type of 'ischaemic heart disease' to place 'Anti-anginal medication' in the 'Drugs used for heart disease' class. In many cases this terminological knowledge is added as a by-product of producing the drug descriptions and so more organization of drug properties may only be available at the later stages of authoring.

The use of bars to signify multiple links at the same level works very well if only a few examples exists per drug description. However we have found that for many descriptions there may be numerous examples. In this case, the number of bars becomes too great for a user to quickly interpret. Experiments are underway to allow the dopamine to temporarily rearrange the structure of the drug description to produce an equivalent description without any multiple links at the same level. For example:

HAS_DRUG_FEATURE indication
FOR prophylaxis
ACTS_ON angina
HAS_FEATURE licensed

Becomes:
HAS_DRUG_FEATURE licensed indication
FOR prophylaxis
ACTS_ON angina

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 beyond the simple structure of the descriptions.

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. Solomon DS, Wroe C, Rogers JE, Rector A. A reference terminology for drugs. Journal of the American Medical Informatics Association 1999(1999 Special Conference Issue).
2. Purves I. Prodigy interim report. Journal of Informatics in Primary Care 1996:2-8.
3. British National Formulary. British Medical Association and the Royal Pharmaceutical Society of Great Britain, 1999.
4. Rogers J, Solomon D, Rector A, Zanstra P. From rubrics to dissections to GRAIL to classifications. Medical Informatics Europe (MIE-97). Thesalonika, Greece: IOS Press, 1997:241-245.
5. Rector A, Rogers J. Ontological issues in using a description logic to represent medical concepts: Experience from GALEN. Methods of Information in Medicine 2000((in press)).
6. Rector A. Coordinating taxonomies: Key to re-usable concept representations. In: Barahona P, Stefanelli M, Wyatt J, eds. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95). Pavia, Italy: Springer, 1995:17-28.
7. Rector AL, Zanstra PE, Solomon WD, et al. Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine. IEEE Transactions on Information Technology in Technology in BioMedicine 1999;2(4):229-242.
8. Solomon D, Rogers J, Rector A, E J van der Haring, Zanstra P. Supporting the use of the GALEN Inermediate Representation. Journal of the American Medical Informatics Association 1998(1998 Special Conference Issue).
9. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The GRAIL concept modelling language for medical terminology. Artificial Intelligence in Medicine 1997;9:139-171.
10. Wroe C, Solomon DS, Rogers JE, Rector A. Inheritance of Drug Information. Submitted to Journal of the American Medical Informatics Association 2000 (2000 Special Conference Issue).