Title:

Design and evaluation of an ontology-based drug application database

Authors:

Christian Senger* MSc[1], Hanna M Seidling* PhD[1,2], Renate Quinzler[1] PhD, Ulf Leser[3] PhD, Walter E Haefeli[1,2] MD

* both authors contributed equally to the work

[1]Department of Clinical Pharmacology and Pharmacoepidemiology, University of Heidelberg, Heidelberg, Germany

[2]Cooperation Unit Clinical Pharmacy, University of Heidelberg, Heidelberg, Germany

[3]Department for Computer Science, Humboldt-Universität zu Berlin, Germany

Author for correspondence:

Walter E. Haefeli, MD

Department of Clinical Pharmacology and Pharmacoepidemiology

University of Heidelberg

Im Neuenheimer Feld 410

69120 Heidelberg, Germany

Phone: +49 6221 56 8740

Fax: +49 6221 56 4642

Email: walter.emil.haefeli@med.uni-heidelberg.de

Keywords: drug information system, ontology, drug administration, electronic decision support

SUMMARY

Objectives: Several recently published cases of preventable adverse drug reactions were associated with flaws in drug application. However, current clinical decision support (CDS) systems do not properly consider drug application issues and thus do not support effective prevention of such medication errors. With the aim to improve CDS in this respect, we developed a comprehensive model precisely describing all aspects of drug application.

Methods: The model consists of (1) a schema comprising all relevant attributes of drug application and (2) an ontology providing a hierarchically structured vocabulary of terms that describe the possible values of the schema's attributes. Finally, medical products were annotated by a semi-automatic term assignment process. For evaluation, we developed an algorithm that uses our model to compute a meaningful similarity between medicinal products with respect to their drug application characteristics.

Results: Our schema consists of 22 attributes. The ontology contains 248 terms, textual descriptions, and synonym lists. More than 58,700 medicinal products were automatically annotated with >386,600 terms. 2,450 drugs were manually reviewed by experts, adding >4,500 terms. The annotation and similarity measure allow for (similarity) searches, clustering, and proper discrimination of drugs with different drug application characteristics. We demonstrated the value of our approach by means of a set of case studies.

Conclusion: Our model enables a detailed description of drug application, allowing for semantically meaningful comparisons of drugs. This is an important prerequisite for improving the ability of CDS systems to prevent prescription errors.

I. BACKGROUND AND INTRODUCTION

In recent years, computerized physician order entry linked to clinical decision support (CDS) systems has become a promising platform to prevent medication errors [1,2]. However, the success of such systems depends on the specificity and clinical relevance of the presented alerts [3]. In many cases, consideration of particularities of drug application will determine the appropriateness of the alerts. For instance, the maximum tolerated daily dose of the antifungal agent amphotericin B strongly depends on the dosage form with 3-4-fold higher dose limits for liposomal compared to aqueous formulations and corresponding dosing errors ignoring these differences have reportedly caused several fatalities [4,5]. CDS systems that aim to warn against prescription of excessive doses must therefore unmistakably refer to the correct upper dose limit and link their knowledge to a very detailed level of the dosage form. Hence, detailed knowledge on dosage forms is mandatory.

This is particularly challenging if active ingredients are available in many different dosage forms (e.g. diclofenac). Referring to the important domain of CDS systems for prevention of drug interactions, the appropriateness of many drug interaction alerts depends on the route of administration. For instance, the antimicrobial ciprofloxacin can be administered by three different routes of application i.e. intravenous, peroral, and topical administration. However, metabolic drug interactions only become relevant for systemically available formulations (i.e. intravenous and peroral) and drug interactions affecting drug absorption actually become manifest only if ciprofloxacin is administered orally [6]. In contrast, systemic drug interactions of topical ciprofloxacin (e.g. administered as eye drops) are not expected. Therefore ignoring the specifics of drug administration will lead to inappropriate interaction alerts whenever a drug may be administered topically and systemically. Similarly, alerting for drug incompatibilities, which

mainly occur during parenteral co-administration, will also require information on administration characteristics. Finally, the route of administration will also determine dose (e.g. for drugs with a high hepatic first-pass) and the potential requirement to adjust doses to renal dysfunction (which is only necessary for systemically available drugs).

Many drug interaction databases do not link their warnings with the drug's availability [7] and may therefore promote over-alerting and alert-fatigue. Moreover, consideration of the route of administration may establish a completely new domain of safety alerts, e.g. preventing intrathecal administration of vincristin, which caused several dozens of deaths [8]. To date, such differentiated knowledge on appropriate as well as prohibited routes of administration is often not available in a structured form, especially if a drug holds a myriad of different routes of administration (e.g. lidocaine).

Last but not least, consideration of drug application is also important for drug switching which has become increasingly important at the interface between different health care sectors [9], in primary care for economic reasons [10], or whenever the change of dosage form or route of administration will determine the therapeutic success.

To guarantee an error-free and effective performance, CDS systems need to be equipped with an unambiguous and precise classification of medicinal products, including dosages, application forms, and routes of administration. Even subtle differences within these attributes need to be explicitly represented and taken into account when comparing and combining different drugs. In order to define appropriate values of the attributes that precisely capture the differences, a controlled vocabulary (ontology) is mandatory. In general, expert knowledge is available in verbal form, free-text, or written semi-formal rules – leading to a "knowledge

acquisition bottleneck" [11]. Thus, ontologies are necessary to allow encoding of medical knowledge and provide machine-interpretable rules for CDS.

We developed a hierarchical model of drug application attributes (schema) together with an ontology defining possible values individual attributes can take. To assess the ability of the model to appropriately describe all marketed drug products, the products were annotated by an automatic approach and a representative cross-sectional sample was annotated by experts. Subsequently, the model's suitability for prevention of drug application errors and for the assessment of the similarity of different drug application forms was assessed. For this purpose, four representative active ingredients referring to 217 different brands were selected, which are known to cause drug application-dependent medication errors or require detailed drug application characteristics for an appropriate evaluation of dosage or drug-drug interactions in CDS systems. Those brands were completely annotated using the controlled vocabulary terms and their similarity based on the annotation was measured.

Aims

We developed a model encompassing attributes and concepts necessary to unambiguously describe all important aspects for safe and secure drug application. All functions of CDS systems that are affected by drug application were to be considered. Moreover, it was important to represent the concepts in such a way that meaningful semantic comparisons between different drugs become possible. In particular, we wanted to support the following tasks:

- searching for drugs with specific drug application characteristics,

- clustering drugs with respect to similarity in drug application,
- linking clusters of similar drugs to specific constraints and rules, and
- searching drugs highly similar to a given drug, where the similarity measure should take into account user-defined constraints (such as "must have" values) and the relationships of annotated concepts in the ontology.

Ontologies and prior work

In this work, we use the term "ontology" to denote a set of concepts covering all important aspects of a domain (here: drug application) and their semantic relationships. Concepts have names, are described by a human readable definition, and may have synonyms. We use the word "term" to denote a name of a concept in the ontology. Such a term may consist of several words.

For our purposes, ontologies serve several functions:

- They provide a standard vocabulary for annotations of biomedical entities, helping to integrate and to search across data sources of different origins and authors [11].
- With ontologies, terms and their domain of application could be specified as exact as possible for the representation of medical knowledge in algorithms and knowledge bases and therefore, sensitivity and specificity of algorithms and knowledge bases can be enhanced [12,13].
- Ontologies enhance search capabilities because the semantic relationships between concepts implicitly define groups. For instance, the ontology may define that `solution`, `emulsion`, and `suspension` are all specializations of the concept

`liquid formulation`. Searching for `liquid formulation` will then also return all drugs annotated with any of these three dosage forms.

- These implicitly defined groups also enable the specification of statements about groups of annotated objects. For instance, one may express that patients on anticoagulants should not be given an intramuscular injection of diclofenac.
- Ontologies enable meaningful similarity searches among drugs. For instance, based on the relationships between concepts in an ontology, an algorithm is able to decide that a capsule is more similar to a tablet than a syrup. Such comparisons are, for instance, an essential prerequisite for drug switching. There are several suggestions on how such a similarity should be computed, ranging from simple term equality over exploitation of the hierarchical relationships between concepts [14] to methods that also employ word frequencies (see Popescu et al. [15] for an example using ICD codes) or the number of possible word meanings [16].
- Finally, ontologies enhance object and concept recognition and the extraction of relationships between objects for information extraction from natural text [17]. They can, for instance, help to automatically extract dosage forms from textual prescriptions.

Prominent examples for ontologies in part covering drugs, drug application, and application units are the Unified Medical Language System (UMLS) [18], the Logical Observation Identifiers Names and Codes (LOINC) [19], and the Systematized Nomenclature of Medicine (SNOMED CT) [20], a health care terminology primarily used to enhance communication and interoperability in electronic health data exchange.

To our knowledge no such comprehensive model to describe drug applications has been published. Ontologies most closely related are RxNorm, a standardized nomenclature for clinical drugs [21], and the less established Prodigy Drug Ontology [22]. RxNorm is used to annotate drugs of the US market with attributes such as `ingredient`, `strength`, `brand name`, `branded ingredient`, or `dose form`. The Prodigy Drug Ontology considers dosage forms, routes, and application devices. However, the models themselves are not fine-grained enough to support CDS systems with respect to drug application. For instance, the attribute `dose form` combines route of administration and dosage form, but does not define whether the drug reaches systemic availability and will thus neither support dose adjustment in renal failure nor drug interaction alerts. Thus, present models are less complete and lack a semantic structure to comprehensively support the numerous tasks of a CDS considering drug application (e.g. similarity calculations). Certain aspects of drug application are also described in ABDATA (ABDATA Pharma-Daten-Service, Eschborn, Germany) [23], which provides detailed information on dosage forms for all drugs marketed in Germany. It distinguishes between presentation form and administration form. However, it does not organize the definitions in a semantic structure and can thus not be used to deduce dependencies or similarities of parameters of drug application relevant for effective and safe use.

In conclusion available ontologies were built for other purposes and thus lacked important aspects with respect to drug application. Instead of refining one of the already existing ontologies with their obvious limitations we therefore decided to build a new model with a strict and convenient schema as a basis of a versatile CDS system. Concurrently we ensured that integration of the terms of our ontology into existing terminologies will be easy to accomplish.

Example drugs

The effectiveness of our model was exemplarily evaluated with a complete annotation of the hospital formulary of the University Hospital of Heidelberg (N=2,450), giving a representative cross-section of all marketed drugs (assuming that about 53% of all described drugs outside a hospital are also found on the hospital fomulary [9]). Furthermore, four frequently used active ingredients were chosen because they represent prominent published examples of drugs that are known

(1) to cause medication errors because of drug application issues (i.e. amphotericin B) [4,5], (2) to trigger over-alerting if drug application characteristics are neglected (i.e. ciprofloxacin) [6], or (3) that are available in many different dosage forms and thus demand a comprehensive and detailed classification of dosage forms, i.e. diclofenac, with 171 drug products with over 30 different (but imprecise and overlapping) terms for application forms as supplied by manufacturers, (May 2009), and (4) that can be administered by many different routes of administration and thus demand a comprehensive classification of routes of administration, i.e. lidocaine with 150 drug products with over 35 different (also vague and overlapping) terms for application forms as supplied by manufacturers, (May 2009).

II. METHODS

Schema development

Previous definitions concerning drug application focused on the dosage form of the drug and its route of administration [21,22,24]. These attributes represent an essential part of but do not sufficiently map all aspects of drug application. On the basis of these attributes a first version of our schema was defined. We then identified further attributes that were indispensable for

evaluation of drug prescriptions and within this process added or modified several attributes, such as the `availability` of drugs (addition), the way in which systemic availability is reached (attribute `absorption`, addition), or the differentiation between `dosage form` and `route of administration` (refinement).

Ontology development

The strategy for identifying concepts followed a computer aided bottom-up approach because the complete (but ambiguous) descriptions of all drug products were available. A manual top-down approach was also applied in a subsequent step to clarify the structure, discard possible redundancies, and extend the structure and vocabulary if concepts were missing. Schema attributes where interconnected by IS-A relationships as well as ontology concepts. Another expression was applied through connections describing the schema attribute(s) to which individual concepts belong to. Hence, the hierarchic level of each concept was easily accessible for similarity calculations and the linkage of knowledge of CDS on different levels of detail. Schema and ontology were constructed and implemented with a tree-like structure in a Relational Database Management System, utilizing interfaces and forms supplied by Microsoft Access and graph visualization tools.

An interdisciplinary team of two pharmacists, two scientists in the area of bio-/medical informatics, and four physicians/medical students screened definitions of drug application published by the authorities. We searched the online available definitions for drug preparations for human use of the US Food and Drug Administration for dosage forms and routes of administration [22,24], the European Pharmacopeia (EuAB) [25], the Standard Terms as they are

published by the European Directorate for the Quality of Medicines [26], German dosage forms as defined by MMI (Medizinische Medien Informations GmbH, Neu-Isenburg, Germany) [27], and ABDATA (ABDATA Pharma-Daten-Service, Eschborn, Germany) [23].

In total, 1123 definitions were collected and analyzed. Note that a definition usually consists of multiple words or phrases and often contains values for different attributes of our schema. Therefore, a definition in general cannot be mapped to a single ontology concept, but first needs to be broken into semantic units. These units were then mapped to terms of the ontology, added as synonyms, or used to extend the ontology. For example, the definition "Powder for suspension for injection" is fragmented into `powder` (referring to `basic presentation form`), `suspension` (referring to `basic form of administration`), and `for injection` (referring to `mode of application`).

Annotation of drug products

We used our ontology to annotate all drugs currently available on the German market. Drugs were annotated in a three-step process (Figure 1). First, basic definitions referring to galenic formulations and route of administration were obtained from MMI. A translation table was compiled, linking one or more concepts to each basic definition. In the second step, drugs were annotated by utilizing the translation table, assigning the concepts of each definition to individual drugs. This process fails whenever the base data do not contain sufficient detail to uniquely identify the corresponding concept of the ontology. Moreover, some aspects of drug application such as information on systemic availability are not covered by traditional definitions of drug application. In a third step we therefore manually completed the annotation for the hospital formulary. Manual curation was supported by an application that provided a user-

friendly graphical interface that also guaranteed the logical dependencies specified in the schema. The experts screened the summary of product characteristics (SPCs) for relevant information and, if information was still lacking, consulted manufacturers for detailed information. Annotation guidelines specifying the selection of appropriate terms were defined and six experts were trained accordingly. For quality assurance, every expert annotated a training set of 15 medicinal products with 159 terms on average. Annotations were stored traceable and all differences were discussed considering their possible influence on the ontology but modifications were not necessary. Experts were trained until full accordance with the respective training set of predefined annotations. Successful annotation of the training set was a prerequisite for working with the real data set. Annotations of the real data were cross-checked by the experts, potential deviations discussed and annotations were adjusted if necessary.

Annotation of drugs with wide and complex ranges of dosage forms

Marketed brands containing diclofenac and lidocaine were used as paradigm drugs to challenge the versatility of the model and to assess the ability to annotate drugs with a wide and complex range of dosage forms or routes of administration.

Comparison of drugs with respect to drug application

Most applications of our ontology will require to reliably identify similar drugs (e.g. for drug switching) or to describe drugs in such detail that information on drug-drug interactions or maximum upper dose limits can be distinguishably and unequivocally linked. We therefore developed a respective method taking the schema and the properties of our ontology into account. Given a query drug, the method returns a list of drugs with the same active ingredients

sorted by similarity to the query drug with respect to drug application. Within the search, the schema is used to determine which values must be compared with each other. Within each attribute, the structure of the ontology attached to this attribute determines the concrete similarity value of annotated drugs.

However, when comparing two drugs certain attributes may need to be treated differently from other attributes. Assume that a solid diclofenac product should be switched to a liquid formulation because of difficulties swallowing (dysphagia). In this situation, it is important to find a drug with a liquid formulation which differs as little as possible from the query drug. Therefore, some characteristics are mandatory (liquid formulation) whereas others (e.g. packaging) are less important for this application. In the following, we use the term ‘soft constraints’ to refer to those attributes whose values should only be as similar as possible, and we use ‘hard constraints’ for conditions on attribute values that must be met.

1. Computation with soft constraints

To compute soft constraints we adapted the Optimistic Genealogy Measure [28]. Trees represent the taxonomy of concepts belonging to a certain attribute of the schema. As an example, the tree for `availability` is shown in Figure 2a with individual nodes representing the terms that can be assigned to a drug. Subtrees contain all nodes between the annotated concept and the root of the tree. The subtree of a drug annotated with the concept `systemical_enteral` is shown in Figure 2b. A drug annotated with `systemical_parenteral` forms the subtree shown in Figure 2c.

The similarity of two drugs is computed by first determining their similarity for every attribute with respect to the lowest common node of both trees. The overall similarity of the

drugs is computed as the average over the similarities of all attributes. Weights for each attribute allow specific attributes to have a greater (or smaller) influence on the similarity of drugs (e.g. systemic/topic availability may be more relevant in drug switching than the packaging of a drug). For the purpose of this study we always set these weights to 1. This similarity measure is asymmetric, i.e. the similarity of drug 1 with drug 2 usually is not the same as the similarity of drug 2 with drug 1. This enables meaningful comparisons of drugs annotated with well defined concepts to those annotated only approximately. For instance, the similarity score of a drug annotated with `systemical_parenteral` (Figure 2b) and one annotated with `systemical` (Figure 2d) will be 0.5, while the reverse comparison will result in a similarity of 1.

2. Computation with hard constraints

Hard constraints are included in our calculations by a function resulting in 1 if all such criteria are matched and 0 in all other cases. The result of calling this function is multiplied with the similarity value to compute the overall result. Accordingly, the overall similarity of two drugs is 0 if any hard criterion is missing.

Similarity searches with annotated drugs

To evaluate functionality and effectiveness of the ontology, the annotations, and the similarity function we conducted similarity searches. We used marketed brands of ciprofloxacin and amphotericin B as paradigm drugs. We applied searches using only soft constraints and subsequently narrowed down the list by adding hard search constraints.

Lists with drug products containing ciprofloxacin and amphotericin B were given to three pharmacists performing a manual similarity ranking with respect to soft constraints. Notice of Table I was given to the pharmacists as similarity criteria. Agreement of the results of automatic ranking and expert ranking was measured using Cohen's weighted kappa coefficient, comparing each expert ranking with the automatic ranking.

III. RESULTS

Our model comprises:

- a schema specifying all relevant attributes of drug application,
- a list of all concepts that can be assigned to the attributes of the schema,
- the semantic relationships between concepts, and
- the description of concrete drug products with the most specific concepts of our ontology for all relevant attributes.

Because the set of attributes describing drug applications has an internal structure, the schema also specifies groups of related attributes. The interplay between schema, ontology, and annotations is depicted in Figure 3.

Schema description

The schema contains 22 attributes and their descriptions (Table I) which are organized in a hierarchical manner (Figure 1). Four main attributes are connected to a general node `drug application`, whose use is purely technical (as a point of entry for structure traversal). They represent the following information:

- `Pharmaceutical form` combines the `presentation form` (`type of presentation form`, `basic form of presentation`, `presentation form`), i.e. the form in which the drug is stored, the `administration form` (`type of administration form`, `basic form of administration`, `administration form`), i.e. the form in which the drug is finally administered, the `drug release`, and the `dosage unit`.
- `drug administration` pools the `site of administration`, defined as "organ" where the drug is supposed to be administered, with the `route of administration` and the `mode of administration`.
- `Absorption` defines whether and how a drug reaches systemic availability.
- `Packaging` specifies the `container` and potential `administration devices` in the packaging.

For each main attribute sub-attributes are defined - e.g. the `type of presentation form` is one aspect of the `pharmaceutical form`. The concrete values of the attributes often imply logical constraints on the values that child nodes may take (not shown graphically). For instance, the `route of administration` imposes restrictions on the `site of administration` (e.g. if `administration site` takes the value `use in the vascular system`, the value `intraocular` is not an option for the `route of administration`).

Ontology description

The final ontology contained 248 terms with descriptions and synonyms. Concepts are connected in taxonomies via `IS-A` relationships e.g. `oily solution|IS-A|solution`. Having concepts at different levels of granularity allows for approximate annotation. For instance, if it is known that a drug's `basic presentation form` is a `tablet` but unknown whether it is a `coated tablet` or an `uncoated tablet`, the more general concept `tablet` should be used.

The relationships between the schema and the ontology ensure the compliance of the concepts with possible constraints implied by the hierarchy of attributes of the schema.

Annotation

The ontology and terms allowed for the approximate automatic annotation of all drug products currently available on the German market (>58.000). The representative subset of the most frequently prescribed drugs (hospital formulary, N=2,450 drug products) was successfully annotated by experts with 29,300 terms to demonstrate the ability of our model for exact annotation of all available drug products. A fraction of 77% of those terms was attached by the automatic annotation step based on the translation table. Correctness of the experts' annotations was tested by annotation of a randomized sample (N=20, 100% term identity, 46 of 255 terms were annotated additionally in the second annotation) by a pharmacist.

Annotation of diclofenac

All 171 products containing diclofenac could be annotated successfully. It required 61 different terms for 17 schema attributes (79 distinct attribute-term combinations) to describe the many different dosage forms and routes of administration. The products were annotated with 8

distinct administration forms and 7 distinct presentation forms encompassing 10 different dosage forms in combination.

Annotation of lidocaine

Lidocaine containing drugs with different dosage forms and/or different application modes (N=31) were successfully annotated with 56 different terms for 14 schema attributes, using 70 different attribute-term combinations. The drugs were annotated with 10 distinct sites of administration, 14 distinct routes of administration, and 9 distinct modes of administration, describing 24 different application forms and/or modes.

Similarity searches

Ciprofloxacin

An aqueous solution for infusion (systemical parenteral availability, ready for use preparation) was used as query drug (Table II). The algorithm found three drugs with identical properties of drug application (score 1.0). Narrowing the search by adding the hard constraint `availability=systemical`, scored "eye drops" 0.0 whereas all other drugs were scored as before. To simulate prevention of drug-drug interactions, the search was further narrowed to drugs with `availability=systemical parenteral` leaving only the three drugs with identical attributes as a result. Experts' ranking was identical, while mean κ of experts and automatic ranking was 0.51. The experts considered the syrup more similar, which was ranked last in the automatic ranking. This was caused by the differences in attaching importance to the attributes `presentation form` and `need for preparation`.

Amphotericin B

The query drug was an aqueous solution represented as a powder to be dissolved before application. Constrained searches were conducted with `type of administration form=liquid formulation` yielding three comparable drugs, whereas restriction to `administration form=aqueous solution` revealed that no other drug than the query drug itself had the desired attribute (Table III). Experts' mean inter-rater agreement was $\kappa = 0.73$ while mean κ of experts and automatic ranking was 0.74 (0.56, 0.82, 0.83).

IV. DISCUSSION

It is a common flaw of current CDS systems that they only incompletely consider drug application characteristics [7]. This is in part caused by the lack of an appropriate model describing pertinent drug properties. In this paper, we present a model that enables the unambiguous characterization of drug application. We believe that such a model will enhance the performance of CDS in the many instances when drug application characteristics are relevant. Key research in the area of medication related CDS has focused on the extension of CDS regarding usability, acquisition of new expert knowledge, and consideration of patient conditions [29,30,31]. The use of ontologies for knowledge acquisition and utilization in CDS knowledge bases is an accepted approach. However, up to now, knowledge is often linked to the active ingredient rather than a specific drug product. Hence, the description of the products has not attracted attention in the literature – even though an improvement of the specificity of warnings and a possible subsequent reduction of alert overriding appears obvious.

Drug switching, a common task for pharmacists, physicians, and also CPOE systems can be facilitated with the proposed model and the demonstrated similarity calculations – supporting manual as well as automated switching. Furthermore, knowledge can be linked to certain drug characteristics in different levels of detail allowing CDS knowledge bases to distinguish between drug products with supposedly small differences in their characteristics (e.g. the occurrence of undesirable effects of intramuscular injections if anticoagulants have been given previously - while intravenous injections are adequate [32]) or substantially different dose requirements of high first-pass compounds if they are given orally as compared to parenteral routes of administration.

To challenge the model we first used an automatic annotation approach and could successfully annotate all available drug products at least approximately. The local hospital formulary as a representative subset of all available drug products was successfully annotated by experts as confirmed by an optimal inter-rater agreement. However, several terms were missing in the first annotation, leading to a less detailed while still correct and useful annotation. The successful annotation of diclofenac and lidocaine products proves the ability of the model to describe drugs with numerous or very complex dosage forms and largely differing administration modes and routes. Thus, we assume that each aspect of drug application for all available drug products could be represented by our model.

To further challenge our model we defined tasks representing important clinical situations in which switching to alternative drugs with similar or distinctly different drug application characteristics is necessary. Brands of amphotericin B and ciprofloxacin are largely differing in drug application forms. Regarding amphotericin B, the essential fact of whether the drug is an

aqueous or liposomal solution can be specified - allowing for appropriate excessive dose warnings. The extent of agreement of expert and calculated ranking confirmed the usability of model and algorithm as an intuitive measure. One expert did not consider the difference in liposomal and aqueous solutions, highlighting the necessity for comprehensive drug application descriptions. Using the example of ciprofloxacin, the model correctly defines the drug's availability and will therefore help to restrict drug-drug interaction alerts to clinically relevant warnings. Moreover, the example of ciprofloxacin illustrates the ability of our model to derive a reasonable and useful scoring even if drugs are annotated only incompletely. A distinctive feature of our scoring scheme is its intentional asymmetry. Drugs that are annotated only approximately are rated relatively low in Table II. However, the inverse comparison would result in a similarity score of 1.0. We believe that this behavior appropriately reflects clinical practice, where specific information usually will be preferred over only approximately described drugs. The ranking, however, could be further improved by applying weights to the similarity measure. For instance, it is often advisable to give a greater weight to the attribute `administration form` and a lower weight to `need for preparation` or `presentation form`. Using such a schema for the examples shown previously, the syrup in Table II and the suspension in Table III would be ranked as more similar to an aqueous solution for infusion than tablets (which are rated higher in the similarity calculations). Clearly, the optimal setting of the different weights depends on the requirements of a specific application. A machine-learning approach with comprehensive training sets related to practice and classified by physicians and pharmacists would be a suitable approach.

Ontologies like RxNorm or the Prodigy Drug Ontology have been developed with the purpose to generally address all medication-related aspects. Thus, while their range is very

broad, they are missing important details in certain domains especially concerning drug application characteristics (e.g. differences between presentation and application form). The presented model exclusively and comprehensively models drug application. The model, annotation, and similarity measure as a whole facilitate drug searches (e.g. in prescription processes) if patients with restrictive conditions like dysphagia need to be treated or switched to more appropriate galenic formulations. Moreover, because the form of administration may also determine maximum tolerated doses, alerts can consider such differences once this information is available in structured format. The different levels of detail and an optional use of the similarity measure allow for clustering of drug products regarding their characteristics and a subsequent linkage of knowledge suitable for use in CDS without the need of an explicit knowledge linkage for each drug product or the most detailed application form (e.g. in the case of dysphagia all solid application forms can be considered at once, instead of considering capsules, tablets, dragees, etc. individually). Note that the model distinguishes between presentation and administration forms specifying for example that effervescent tablets do not have a solid administration form.

Hence, the model and the attached similarity measure eliminate several shortcomings of current drug information encoding systems:

- First, the ontology terms and their description allow for an unambiguous communication of drug application characteristics.
- Second, the level of detail in the annotation of drugs enables a new type of safety alerts in CDS systems that are urgently needed to prevent potentially life-threatening medication errors. If for example a physician chooses an inappropriate route of

administration for a medicinal product (e.g. intrathecal administration of vincristine) an alert may be issued.

- Third, also the specificity of many alerts of CDS systems can be substantially enhanced if drug application characteristics are considered. As an example, drug-drug interaction alerts are usually only relevant if the drugs are systemically available and do not occur if either drug's availability is topical. The quality of alerts (e.g. the severity) can be adjusted based on the detailed description of drug application (e.g. major interaction if intravenous β-adrenoceptor antagonists are co-administered with intravenously administered calcium channel blockers compared to an only moderate interaction with oral verapamil).

The model has been established and integrated into the CDS system at the University Hospital of Heidelberg for 1½ years and confirmed its quality and performance in several knowledge base fields including drug-drug interactions. While we believe in the advantages of the model, its benefits have to be proven in a prospective study.

During development and curation of the knowledge bases, scientists give permanent feedback for evaluation and extension of the model. The model is updated every second week with newly available market data. Provision of access to the model for data suppliers and future research projects is planned for the near future.

The current model also has limitations. For instance, we did not include the dosage of medicinal products while other systems, including RxNorm, do model drug dosage. However, drug dosage is rather a link to other medicinal ontologies because drug dosage does not only depend on drug application characteristics but also on the indication of the drug and the patient's individual condition (e.g. kidney function), and co-medication. Properly reflecting these

dependencies therefore requires knowledge of the patient's co-morbidities i.e. links to ICD-10 codes and lab values. Moreover, divisibility of tablets was studied previously [33] and can be attached to the model for all drugs with `basic form of administration=tablet`. Only drugs marketed in Germany were annotated, which is one of the largest drug markets of the world. However, annotation can be easily extended to other drugs if the respective information on drug application characteristics is extracted from the label. Finally, annotation of medicinal products requires a frequent update process to ensure high quality and current data. With the two-weekly update, a screening of the SPCs that have been updated by the manufacturer during the preceding two weeks is necessary. Obviously such an update process would be facilitated if the regulatory authorities would ask for such information as a prerequisite for marketing authorization and if it would thus be provided in structured format by the pharmaceutical manufacturer (e.g. as a structured electronic SPC).

V. CONCLUSION

The presented model closes a gap in current CDS systems used in pharmacotherapy by offering a comprehensive and detailed characterization of drug application, which is a prerequisite for highly-specific alerts and drug switching.

References

1. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety. A systematic review. Arch Intern Med. 2003;163:1409-16.

2. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. BMJ. 2005;330:765-72.

3. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc. 2006;13:138-47.

4. Mohr JF, Hall AC, Ericsson CD, Ostrosky-Zeichner L. Fatal amphotericin B overdose due to administration of nonlipid formulation instead of lipid formulation. Pharmacotherapy. 2005;25:426-8.

5. Burke D, Lal R, Finkel KW, Samuels J, Foringer JR. Acute amphotericin B overdose. Ann Pharmacother. 2006;40:2254-9.

6. Neuhofel AL, Wilton JH, Victory JM, Hejmanowsk LG, Amsden GW. Lack of bioequivalence of ciprofloxacin when administered with calcium-fortified orange juice: a new twist on an old interaction. J Clin Pharmacol. 2002;42:461-6.

7. Payne TH, Nichol P, Hoey P, Savarino J. Characteristics and override rates of order checks in a practitioner order entry system. Amia Symp Proc. 2002:602-6.

8. Dettmeyer R, Driever F, Becker A, Wiestler OD, Madea B. Fatal myeloencephalopathy due to accidental intrathecal vincristin administration: a report of two cases. Forensic Sci Int. 2001;122:60-4.

9. Walk SU, Bertsche T, Kaltschmidt J, Pruszydlo MG, Hoppe-Tichy T, Walter-Sack I, et al. Rule-based standardised switching of drugs at the interface between primary and tertiary care. Eur J Clin Pharmacol. 2008;64:319-27.

10. Andersson KA, Petzold MG, Allebeck P, Carlsten A. Influence of mandatory generic substitution on pharmaceutical sales patterns: a national study over five years. BMC Health Serv Res. 2008;8:50.

11. Kawazoe Y, Ohe K. An ontology-based mediator of clinical information for decision support systems: a prototype of a clinical alert system for prescription. Methods Inf Med. 2008;47:549-59.

11. The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Research. 2001;11:1425-33.

12. Liu JH, Milstein C, Séné B, Venot A. Object-oriented modeling and terminologies for drug contraindications. Methods Inf Med. 1998;37:45-52.

13. Cimino JJ. The concepts of language and the language of concepts. Methods Inf Med. 1998;37: 311.

14. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22:1600-7.

15. Popescu M, Arthur G. OntoQuest: A physician decision support system based on ontological queries of the hospital database. Amia Symp Proc. 2006;639.

16. Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. Proceedings of International Conference Research on Computational Linguistics (ROCLING X). 1997.

17. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform. 2008:67-79.

18. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nuc Acids Res. 2004;32:267-70.

19. Khan AN, Griffith SP, Moore C, Russell D, Rosario AC Jr, Bertolli J. Standardizing laboratory data by mapping to LOINC. J Am Med Inform Assoc. 2006;13:353-5.

20. International Health Terminology Standards Development Organisation (IHTSDO). Systematised Nomenclature of Medicine – Clinical Terms (SNOMED CT). http://www.ihtsdo.org. Accessed Mar 24th, 2010.

21. Lui S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: Prescription for electronic drug information exchange. IT Professional. 2005;7:17-23.

22. Solomon WD, Wroe CJ, Rector AL, Rogers JE, Fistein JL, Johnson P. A reference terminology for drugs. Proc AMIA Symp. 1999;152-5.

23. ABDATA - Pharma-Daten-Service. Abda-Artikelstamm. Available from: http://www.wuv-gmbh.de/de/pages/abdata/artikelstamm.php. Accessed March 27th, 2010.

24. FDA, CDER. Data standards manual for routes of administration. Available at: http://www.fda.gov/cder/dsm/DRG/drg00301.htm . Accessed December 4th, 2008.

25. Council of Europe. The European Pharmacopoeia, 6th ed. 2007.

26. Council of Europe.Standard Terms. Pharmaceutical dosage forms. Routes of administration. Containers. European Directorate for the Quality of Medicines, 5th ed. 2004.

27. Medizinische Medien Informations GmbH MMI. Gelbe Liste Pharmindex. Available from: http://www.gelbe-liste.de/. Accessed March 27th, 2010.

28. Ganesan P, Garcia-Molina H, Widom J. Exploiting hierarchical domain structure to compute similarity. ACM Transitions on Information Systems. 2003;21:64-93.

29. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc. 2007;14:29-40.

30. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, Campbell E, Bates DW. Grand challenges in clinical decision support. J Biomed Inform. 2008;41:387-92.

31. Seidling HM, Schmitt SPW, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, Bertsche T, Walter-Sack I, Haefeli WE. Patient-specific electronic decision support reduces prescription of excessive doses. Qual Saf Health Care 2010 (in press).

32. Marder VJ. The use of thrombolytic agents: choice of patient, drug administration, laboratory monitoring. Ann Intern Med. 1979;90:802-8.

33. Quinzler R, Gasse C, Schneider A, Kaufmann-Kolle P, Szecsenyi J, Haefeli WE. The frequency of inappropriate tablet splitting in primary care. Eur J Clin Pharmacol. 2006;62:1065-73.

Tables

Table I

Definition of schema attributes. For illustration, the respective ontology terms are shown for an effervescent tablet with systemic availability.

Schema attribute	**Description**
Drug application	General node, point of entry for structure traversal
Pharmaceutical form	The pharmaceutical form is the combination of the form in which a pharmaceutical product is presented by the manufacturer (form of presentation) and the form in which it is administered, including the physical form (form of administration)
Type of presentation form	General type of galenic formulation (e.g. solid, semisolid, liquid formulation) presented by the manufacturer e.g. `solid formulation`
Basic form of presentation	Basic galenic formulation presented by the manufacturer e.g. `tablet`
Presentation form	galenic formulation presented by the manufacturer e.g. `effervescent tablet`
Type of administration form	General type of galenic formulation (e.g. solid, semisolid, liquid formulation) as which the drug is administered e.g. `liquid formulation`
Basic form of administration	Basic galenic formulation as which the drug is administered e.g. `solution`
Administration form	Galenic formulation as which the drug is administered

	e.g. `aqueous solution`
Drug release	Mode and timing of drug release from the administration form after application e.g. `immediate`
Technique of drug release	Galenic technique to modify drug release (e.g. coating, matrix) e.g. `n.a.`
Dosage unit	Unit in which the drug is usually dosed (e.g. tablet, drop, ml) e.g. `tablet`
Drug administration	**Site, route, and mode used to administer the drug to the patient**
Site of administration	Indicates the organ, tissue, or superordinate system on which, through which, or into which the drug is to be introduced. e.g. `oral application for uptake`
Route of administration	Indicates the site of the body on which, through which, or into which the drug is to be introduced e.g. `peroral`
Mode of administration	Indicates how (by which technique) a drug is administered e.g. `to be swallowed`
Preparation	Specifies whether a drug needs to be prepared before administration e.g. `to be dissolved`
Diluent available	Indicates whether a diluent is available in the packaging e.g. `no`
Absorption	**Specifies whether and how the drug is absorbed by the organism**
Availability	Specifies whether a drug reaches the systemic circulation or whether

	its distribution is confined to a specific area e.g. `systemical`
Availability detailed	Specifies how systemic availability is reached (via enteral or parenteral absorption) e.g. `systemical_enteral`
Packaging	**Specifies the packaging of the presentation form**
Container	Immediate packaging of the presentation form e.g. `tube`
Administration device	Indicates whether the packaging contains a device to facilitate the administration e.g. `no`

Table II

Similarity values of medicinal products containing ciprofloxacin. Query drug is an aqueous solution for infusion without the need of preparation.

Application characteristics of paradigm brands	**Similarity value**	**Constrained similarity value**	
		availability= systemical	availability= systemical_ parenteral
aqueous solution for infusion (systemical parenteral), no need for preparation	**Query drug**		
aqueous solution for infusion (systemical parenteral), no need for preparation	1.0	**1.0**	**1.0**
aqueous solution to be dropped on the conjunctiva (topical), no need for preparation	0.5	**0.0**	**0.0**
solution for infusion (preparation, and kind of solution undefined)	0.47	**0.47**	**0.0**
coated tablet for swallowing (systemical enteral), no need for preparation	0.25	**0.25**	**0.0**
syrup for swallowing (systemical enteral), represented as coated granules which have to be suspended	0.14	**0.14**	**0.0**

Table III

Similarity values of medicinal products containing amphotericin B. Query drug is an aqueous solution for infusion.

Application characteristics of paradigm brands	**Similarity value**	**Constrained similarity value**	
		type of administration= liquid formulation	administration form= aqueous solution
aqueous solution for infusion, represented as powder that needs to be solved	**Query drug**		
liposomal solution for infusion, represented as lyophilisate that needs to be mixed	0.60	**0.6**	**0.0**
liposomal solution, no need for preparation	0.38	**0.38**	**0.0**
uncoated tablet for swallowing (topical)	0.07	**0.0**	**0.0**
uncoated tablet for sucking (topical)	0.07	**0.0**	**0.0**
suspension for swallowing (topical)	0.05	**0.05**	**0.0**

Figures

Figure 1

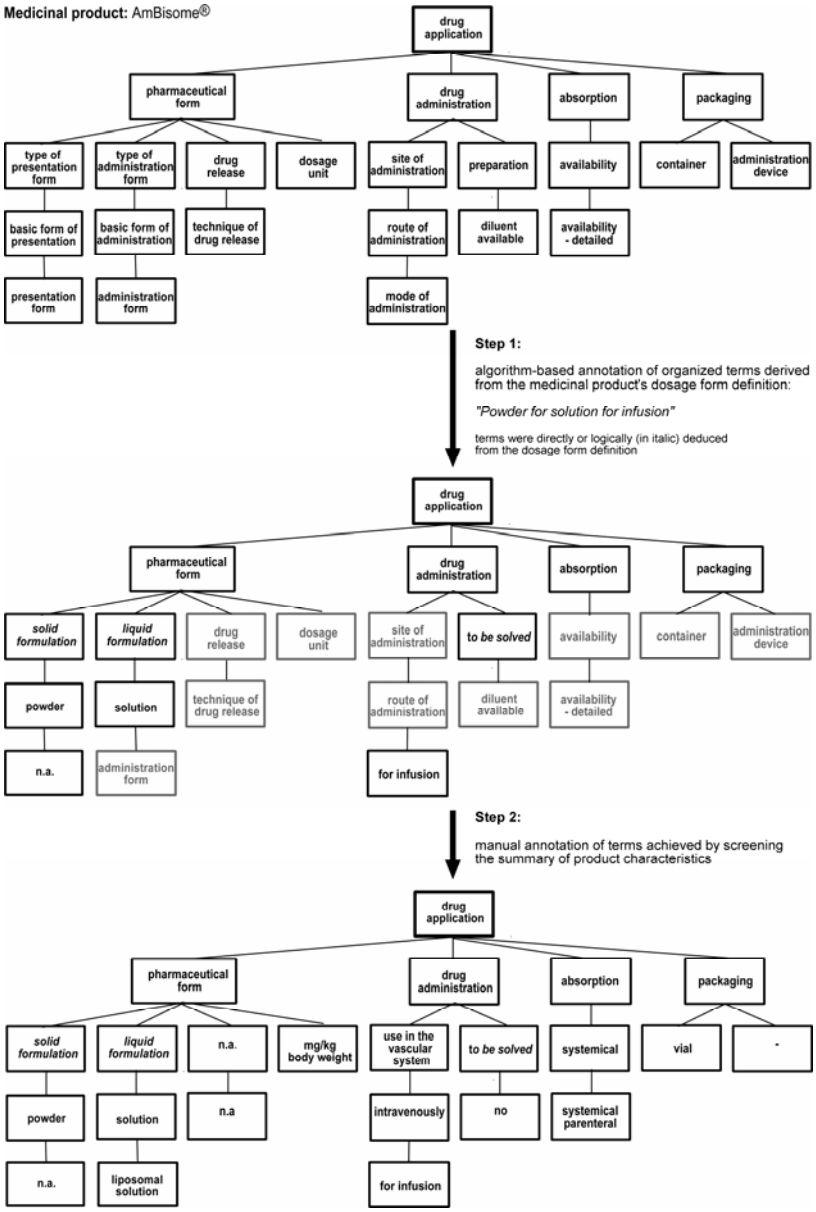


Figure 2

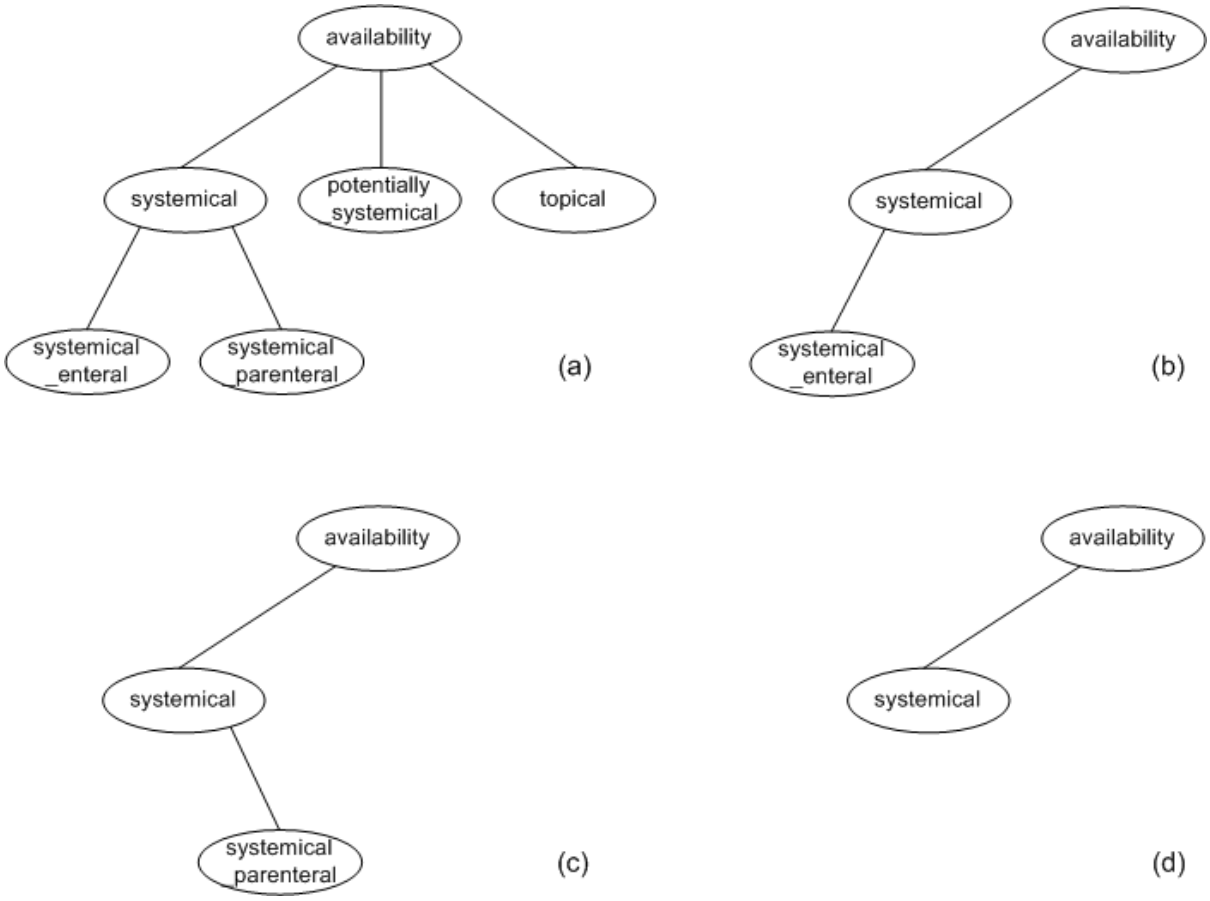
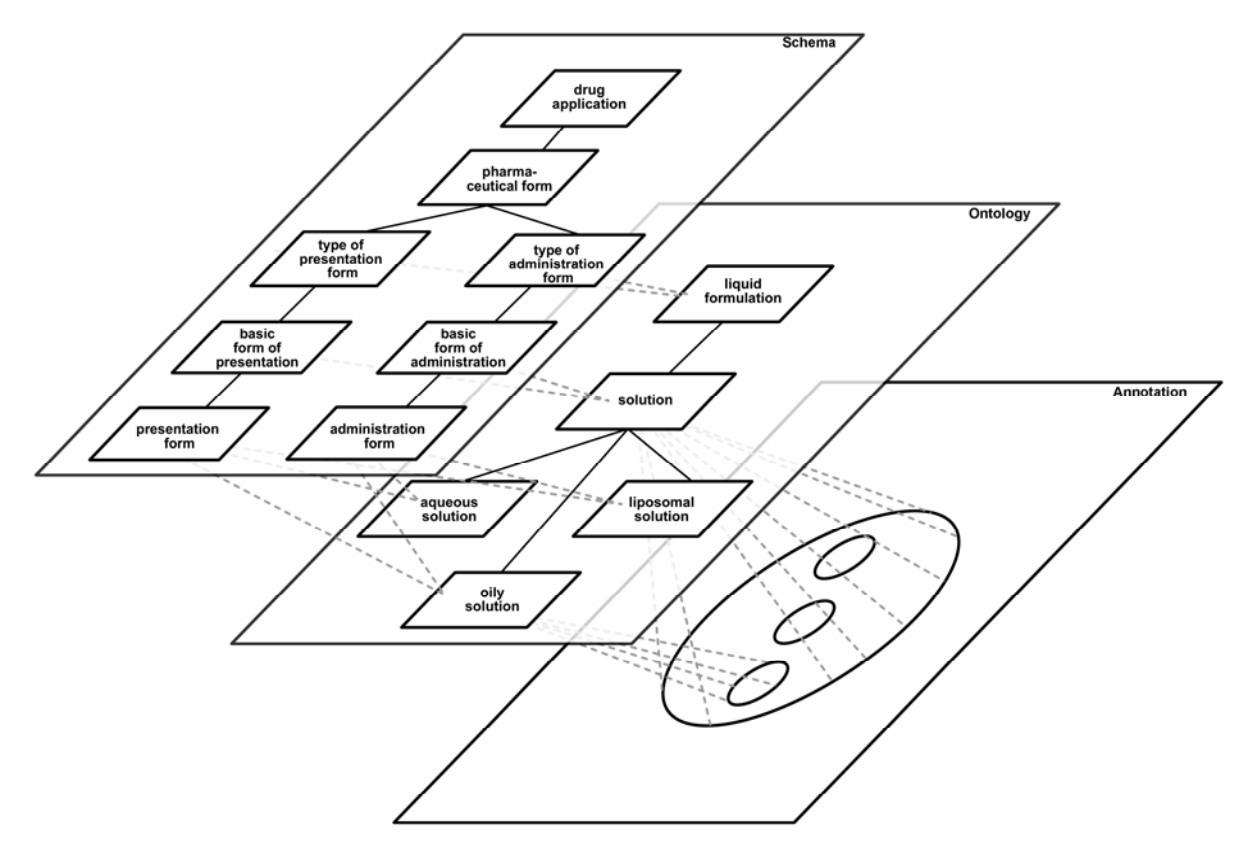

availability
systemical
potentially_systemical
topical
systemical_enteral
systemical_parenteral
(a)
availability
systemical
systemical_enteral
(b)
availability
systemical
systemical_parenteral
(c)
availability
systemical
(d)

Figure 3


Schema
drug application
pharma-ceutical form
type of presentation form
type of administration form
basic form of presentation
basic form of administration
presentation form
administration form
Ontology
liquid formulation
solution
aqueous solution
liposomal solution
oily solution
Annotation

Legends to figures

Figure 1. Schema of drug application characteristics with defined concepts in a hierarchic structure (top) and annotation of medicinal products using the example of a liposomal solution for infusion of amphotericin B. The first step in the annotation process (generation of a translation table) is not included.

Figure 2. (Induced) trees for similarity calculations of drugs annotated with drug application characteristics. (a) Tree related to `availability` with all possible characteristics. (b) Tree induced by a drug with `systemical_enteral` availability. (c) Tree induced by a drug with `systemical_parenteral` availability. (d) Tree induced by an approximate annotated drug with `systemical` availability.

Figure 3. Schematic structure of the model of drug application. Displayed are the relationships of the schema, the ontology, and the drug products using the example of the metaconcepts “pharmaceutical form”. The formal definition of the schema attributes, the inter-schema relationships, and relationships of the concepts allows CDS to link knowledge explicitly to different levels of detail.