Drug-drug Interaction Discovery Using Abstraction Networks for "National Drug File – Reference Terminology" Chemical Ingredients

Christopher Ochs, MS[1], Ling Zheng, MS[1], Huanying Gu, PhD[2], Yehoshua Perl, PhD[1], James Geller, PhD[1], Joan Kapusnik-Uner, PharmD[3], Aleksandr Zakharchenko, BS[1]
[1]New Jersey Institute of Technology, Newark, NJ; [2]New York Institute of Technology, New York, NY; [3]First Databank, Inc., San Francisco, CA

Abstract

The National Drug File – Reference Terminology (NDF-RT) is a large and complex drug terminology. NDF-RT provides important information about clinical drugs, e.g., their chemical ingredients, mechanisms of action, dosage form and physiological effects. Within NDF-RT such information is represented using tens of thousands of roles. It is difficult to comprehend large, complex terminologies like NDF-RT. In previous studies, we introduced abstraction networks to summarize the content and structure of terminologies. In this paper, we introduce the Ingredient Abstraction Network to summarize NDF-*RT's* Chemical Ingredients and their associated drugs. Additionally, we introduce the Aggregate Ingredient Abstraction Network, for controlling the granularity of summarization provided by the Ingredient Abstraction Network. The Ingredient Abstraction Network is used to support the discovery of new candidate drug-drug interactions (DDIs) not appearing in First Databank, Inc.*'s DDI* knowledgebase.

Introduction

We present a new method for the discovery of candidate drug-drug interactions (DDIs). Conceptually, this method is based on comparing a large commercial knowledgebase of DDIs with small groups of drugs where all members of one group contain similar ingredients. These groups of drugs were derived from an independent drug terminology, the National Drug File – Reference Terminology (NDF-RT) [1]. The challenge is that the NDF-RT in its source format does not provide appropriate groups of drugs. Finding such groups by hand is difficult, because the NDF-RT is composed of approximately 43,000 concepts connected by 67,000 IS-A roles and 73,000 other roles, e.g., has ingredient and has mechanism of action. Thus, we are using algorithms, based on a well-developed body of work on abstraction networks (AbNs) [2], to derive small groups of similar drugs. Intuitively, if most but not all members of such a group have known drug-drug interactions according to a DDI knowledgebase, then the remaining group members should be reviewed by an expert to determine why they don't have (known) DDIs. The challenge is that none of our previously developed abstraction networks could be applied to the NDF-RT. In this paper, we present a new AbN, expressly developed for the NDF-RT and for the purpose of discovering candidate DDIs.

NDF-RT uses a description logic-based concept model to define drugs according to various aspects, e.g., active chemical ingredient, mechanism of action, physiologic effect, therapeutic intent, and dosage form. Each aspect is represented by a separate concept hierarchy and the aspects of each drug are expressed via roles. For example, the drug concept Aspirin/Caffeine has two has ingredient roles, to Aspirin and to Caffeine.

Abstraction networks (AbNs) [2] are compact, visual terminology summaries, based on grouping "similar" terminology elements together into groups. (The meaning of "similarity" is terminology-dependent.) One type of AbN that was previously applied to description logic-based terminologies is the partial-area taxonomy [3], thus one would expect that this AbN could be applied to the NDF-RT. However, due to NDF-RT's highly specialized structure, our previously developed methods for deriving AbNs are not applicable to it, because the majority of NDF-RT's hierarchies have no roles emanating from their concepts. The existence of such roles is the basis for our previously developed AbNs for description logic-based terminologies.

In this paper, we describe the derivation of a new kind of AbN called an Ingredient Abstraction Network (IAbN), which summarizes NDF-RT's chemical ingredients and their associated drug concepts. The NDF-RT distinguishes between different dosage forms of the same drug. However, for the purpose of the discovery of potential DDIs we can safely ignore such distinctions in our algorithms. The resulting IAbN turned out to be not sufficiently compact, which is a desirable feature. Thus, we developed a secondary abstraction mechanism ("Aggregate IAbN") that creates an IAbN with fewer groups and each group stands for a collection of groups from a "complete" IAbN (i.e., a non-aggregate IAbN). The Aggregate IAbN derivation algorithm is parameterized to allow users control over the degree of summarization that is achieved.

Background

NDF-RT is a formal representation of the VHA National Drug File (NDF) [4], which is a drug classification hierarchy used to group orderable drug products into one of 579 drug classes. NDF is used to support VHA's clinical applications. NDF-RT's Mechanism of Action, Physiologic Effect and Chemical Ingredients hierarchies were created by matching VHA drug ingredient names to terms from the National Library of Medicine (NLM)'s Medical Subject Headings (MeSH) [5]. Specifically, the Chemical Ingredients (CI) hierarchy was derived from MeSH's Chemicals and Drugs Category. The Mechanism of Action and Physiologic Effect hierarchies were also initially created based on MeSH [6] and then extended. NDF-RT organizes concepts around the Pharmaceutical Preparations (PP) hierarchy, the largest one in NDF-RT, with 25,093 concepts (July 2014 version). Besides IS-A roles, PP concepts can have roles to concepts in other hierarchies to define drugs according to their various aspects. For example, the drug concept Aspirin in PP has the role has ingredient to Aspirin in CI, the second largest hierarchy in NDF-RT with 10,118 concepts.

An abstraction network [2] (AbN) is defined as a compact network of nodes that summarize groups of "similar" concepts in a terminology to capture its "big picture." The definition of "similar" is based on the type of terminology and the specific kind of abstraction network. AbN nodes are organized in a hierarchy derived from the hierarchical IS-A links of the terminology. In previous work, we have derived various [3, 7-10] kinds of abstraction networks for many different terminology systems, e.g., the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) [11], the National Cancer Institute thesaurus (NCIt) [12], and the Gene Ontology (GO) [13]. These AbNs were shown successful in supporting the identification of concepts with a high likelihood of errors. For a review of different kinds of AbNs and their properties, see Halper et al. [2].

Extensive research has been done on NDF-RT, e.g., on its content coverage, the adequacy of representation, drug normalization and classification. Rosenbloom et al. [14] investigated the adequacy of the representation of the Physiologic Effect hierarchy. The results suggested that the concepts in the Physiologic Effect hierarchy are appropriate for medications. Carter et al. [15] studied drug class names from three sources to understand how drugs were classified, and evaluated NDF-RT's semantic coverage. They found that NDF-RT can cover more than 90% NDF drug categories. Pathak et al. [16] evaluated the applicability of RxNorm [17] and NDF-RT for classification of medication data extracted from electronic health records (EHRs). Their study demonstrated that the two terminologies can be used together for drug classification.

Methods

Definition: An Ingredient Abstraction Network (IAbN) is an AbN where the nodes summarize (1) the ingredients in the Chemical Ingredient hierarchy **and** (2) those drug concepts in the PP hierarchy that have no dosage information but that do have at least one has ingredient role to a drug ingredient in the Chemical Ingredient hierarchy.

Refer to Figure 1 for the following definitions that will be used throughout the whole paper. We distinguish between four types of concepts in the Chemical Ingredients (CI) hierarchy. (1) A drug ingredient in CI is the target of has ingredient role(s) from concepts in the Pharmaceutical Preparations (PP) hierarchy. Drug ingredients are chemical ingredients that are used in prescription drugs. (2) A classification ingredient in CI is a concept that "organizes" other drug ingredients. It has drug ingredients as children. (3) A dual use ingredient in CI is both a drug ingredient and a classification ingredient. Note that this is "dual use" in the terminology, not for prescription. A classification ingredient that is not also a drug ingredient is called a strict classification ingredient. (4) Some concepts in CI are neither a drug ingredient nor a classification ingredient. No specific name was chosen for such concepts, as they are not used in this work. The right side of Figure 1(a) illustrates these definitions for an excerpt of 12 CI concepts.

The design of an AbN for the CI hierarchy poses a challenge for several reasons: (1) A lack of roles emanating from CI concepts prevents the derivation of AbNs called partial-area taxonomies [3]. (2) The need to distinguish between drug ingredients and classification ingredients is further complicated by the dual use of many CI concepts. (3) There is a need to summarize the drug concepts, which in NDF-RT are parts of the PP hierarchy, according to their ingredient concepts in CI. (4) To obtain a "big picture" of the classification ingredients, one needs to identify distinct groups of drugs and organize them in a way that supports DDI discovery. (5) A method is needed that allows a user to control the granularity of summarization, so that s/he is not overwhelmed.

IAbN derivation begins with identifying all of the drug concepts in the PP hierarchy with a has ingredient role but with no has_DoseForm role. PP concepts with dosage information are ignored since an ancestor concept, typically a parent (a PP generic drug ingredient) introduces the has ingredient role, which is inherited to such concepts. All of the PP concepts in Figure 1(a), except Pharmaceutical Preparations, have one has ingredient role to a concept in

the CI hierarchy. Different drug concepts in the PP hierarchy can have a has ingredient role to the same CI concept, e.g., both Aspirin and Acetylcilate sodium have the ingredient Aspirin. PP concepts may also have multiple has ingredient roles, e.g., Aspirin/Caffeine has has ingredient roles to both Aspirin and Caffeine.
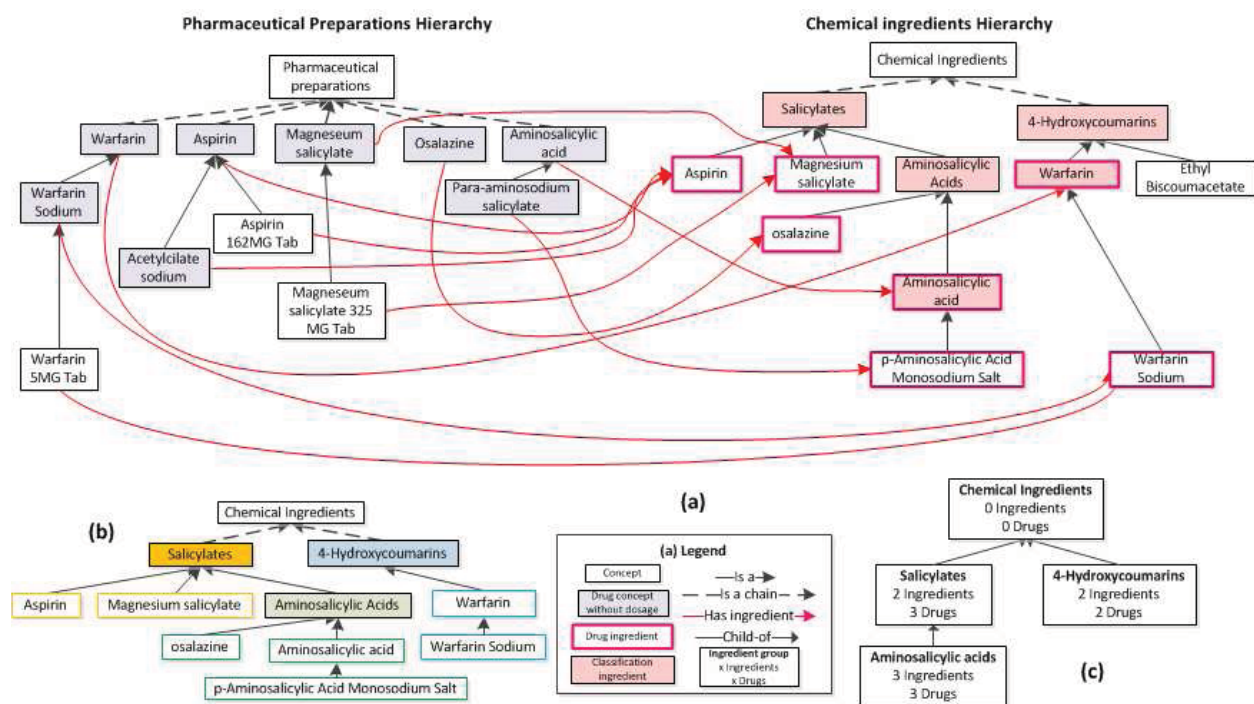

Figure 1. (a) An excerpt of concepts from NDF-RT's Pharmaceutical Preparations (PP) and Chemical Ingredients (CI) hierarchies. On the left, drug concepts in the PP hierarchy with no dosage information have a shaded background. On the right, seven drug ingredients are outlined in pink and five classification ingredients have a pink background. Two concepts, Aminosalicylic acid and Warfarin, are both drug ingredients and classification ingredients, i.e., they are dual use ingredients. Ethyl Biscoumacetate is neither a drug ingredient nor a classification ingredient. **(b)** CI revisited: Drug ingredients (not shaded) and their lowest common ancestor classification ingredients (shaded). Each drug ingredient is color-framed according to its lowest common ancestor classification ingredient. **(c)** The final IAbN for the excerpt of Figure 1(a). Ingredient groups are shown as white boxes that are labeled with the name of the lowest common ancestor from 1(b). Also shown are the total number of ingredient concepts summarized by the group, and the total number of drug concepts [with no dosage information in the PP hierarchy] with has ingredient roles pointing to the CI hierarchy. Child-of links between ingredient groups are shown as upward directed arrows.

Next, drug ingredients (see definition above) are identified by collecting the target concepts of all the has ingredient roles. Classification ingredients (again, see definition above) are identified by analyzing the parent concept(s) of each drug ingredient. Additionally, for each drug ingredient, the lowest ancestor(s) that are a strict classification ingredient(s) are identified, with the intention of finding lowest common ancestor classification ingredients for groups of drug ingredients. For example, for the Aspirin CI concept, the lowest ancestor that is a strict classification ingredient is Salicylates. Salicylates is the lowest common ancestor for Aspirin **and** Magnesium salicylate. For Warfarin Sodium its parent concept, Warfarin, is a classification ingredient but it is also a drug ingredient (i.e., it is dual use). Thus, the lowest ancestor of Warfarin Sodium that is a strict classification ingredient is Warfarin's parent, 4-Hydrocourmarins. Many CI hierarchy concepts have multiple parents, thus, a given drug ingredient may have more than one lowest ancestor which is a strict classification ingredient.

Drug ingredients are grouped together according to their lowest common ancestor(s) that are strict classification ingredients. For example, Aspirin and Magnesium salicylate both share Salicylates as lowest common ancestor. Similarly, Warfarin and Warfarin Sodium share 4-Hydrocourmarins as a lowest common ancestor. Figure 1(b) models the right side of 1(a) and shows the drug ingredient groups induced by the lowest common ancestors.

Next, each strict classification ingredient is made into a root for its ingredient group. Thus Salicylates becomes the root of the group with Aspirin and Magnesium salicylate in it. The CI root concept, Chemical Ingredients, is also a root. Figure 1(c) shows how roots stand in for their groups. The line "2 Ingredients" under Salicylates indicates how much information is summarized. Ingredient groups are not disjoint; drug ingredients with multiple parents may be summarized by multiple ingredient groups. With this we have achieved a summary of the "right side" (the Chemical Ingredients) of Figure 1(a). In the next step, we include information from the left (PP) side into Figure 1(c).

For each ingredient group, the PP drug concepts that have a has ingredient role to a drug ingredient in the ingredient group are identified. For example, the Aspirin and Acetylcilate Sodium drug concepts in PP both have Aspirin in CI as the target of their has ingredient roles. The Aspirin drug ingredient belongs to the Salicylates ingredient group, thus, the Aspirin and Acetylcilate drug concepts from PP are also summarized by the Salicylates ingredient group. This is expressed by the line "3 Drugs" under Salicylates in 1(c). (The third is Magnesium salicylate). Since ingredients may belong to multiple ingredient groups, a given PP drug concept may be in multiple ingredient groups.

Within the IAbN, ingredient groups are organized into a hierarchy according to child-of links derived from the underlying IS-A hierarchy. An ingredient group A is a child-of another ingredient group B if A's root has B's root as an ancestor and there are no other roots on any path from A's root to B's root. An ingredient group may be child-of multiple ingredient groups. In the visualization of an IAbN it is necessary to organize the ingredient groups in a way that helps the summary reflect the "big picture." Thus, ingredient groups are organized into color coded levels according to the length of the longest child-of path to the root ingredient group (Chemical Ingredients). Figure 1(c) shows the IAbN derived from the NDF-RT excerpt in Figure 1(a).

We note that there are CI concepts that are neither drug ingredients nor classification ingredients, e.g., Ethyl Biscoumacetate. (See case (4) in the above definitions and Figure 1(a).) This occurs when the drug ingredient is modeled in CI but no PP drug concept has a has ingredient role to this drug ingredient. For the current research, such concepts are not summarized by any ingredient group and are not considered part of the IAbN. In the Discussion we propose methods for extending the IAbN to include these concepts. Additionally, we note that drug concepts with dosage information are not associated with an ingredient group. However, the ingredient group(s) of these drug concepts can be identified via their parents, which are grouped into at least one ingredient group.

Aggregate Ingredient Abstraction Network (Aggregate IAbN)

One significant issue we encountered when deriving the IAbN for the complete CI hierarchy ("complete IAbN") was its large size. While the complete IAbN is significantly smaller and less complex than the underlying CI hierarchy, there are still too many nodes, many of which "summarize" only one drug ingredient. To improve the efficacy of the IAbN to function as a summary, we now introduce a parametric method for controlling the granularity of summarization. This secondary summarization approach is based on the following heuristic: An ingredient group that summarizes a relatively large number of drug ingredients is more important within the CI hierarchy than an ingredient group that summarizes relatively few drug ingredients.

However, controlling the granularity of summarization provided by the IAbN requires more than just hiding small ingredient groups. Simply hiding these groups leads to a loss of information and inconsistencies in the child-of hierarchy. To obtain a more compact secondary summary of the CI hierarchy's content we have developed a parametric IAbN called an Aggregate Ingredient Abstraction Network ("Aggregate IAbN") that aggregates "small" ingredient groups into their larger direct ancestor ingredient groups.

Given a bound b (a natural number), an Aggregate IAbN is derived in the following way. Starting from the root ingredient group, Chemical Ingredients, the hierarchy of ingredient groups is traversed downwards using a Topological Sort; that means, an ingredient group is processed only after all of its parent ingredient groups have been processed. We define a) aggregate ingredient groups, b) removed ingredient groups, and c) regular ingredient groups. A removed ingredient group is an ingredient group in the complete IAbN with fewer than b ingredients. The root ingredient group, Chemical Ingredients, is by definition **not** a removed ingredient group. Approximately speaking, an aggregate ingredient group is an ingredient group that summarizes itself and one or more removed ingredient groups. A regular ingredient group is not changed when going from a complete IAbN to an Aggregate IAbN. Thus the Aggregate IAbN consists of aggregate ingredient groups and regular ingredient groups.

More precisely, in an Aggregate IAbN, a removed ingredient group i is included into an aggregate ingredient group a if i is a descendant of a in the complete IAbN and there is no other aggregate ingredient group on any child-of path from i to a. Removed ingredient groups are "hidden" in the Aggregate IAbN. Consider the IAbN excerpt in Figure 2 and a bound b=10. Chemical Ingredients has two child ingredient groups with fewer than ten drug ingredients

(Glycosides and Heterocyclic Compounds) and several further descendants with fewer than ten drug ingredients (e.g., Glucosides and Piperidones). Some of these further descendant ingredient groups may be direct descendants of other ingredient groups that have more than ten drug ingredients, e.g., Piperidones is child-of Piperidines, which summarizes 47 ingredients.

Consider the two removed ingredient groups that are children of Chemical Ingredients in the IAbN of Figure 2(a), with b=10: Glycosides and Heterocyclic Compounds. There are no intermediate aggregate ingredient groups, so these children are included into Chemical Ingredients in the Aggregate IAbN of Figure 2(b). Furthermore, the removed ingredient group Glucosides is a child of Glycosides. The closest aggregate ingredient group above Glucosides that is not a removed ingredient group, via the path through Glycosides, is Chemical Ingredients. Thus, Glucosides is included into the Chemical Ingredients aggregate ingredient group as well. Note: I=14 in Chemical Ingredients in Figure 1(b), summarizing the removed ingredient groups, while I=3 in Chemical Ingredients in 1(a).
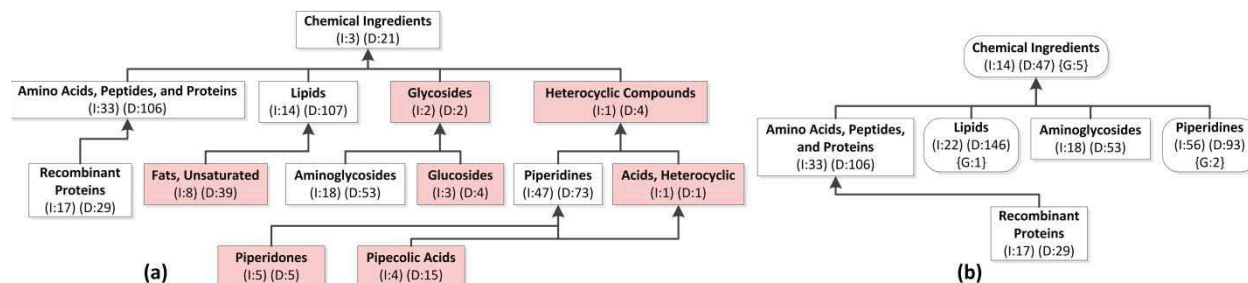

Figure 2. (a) An excerpt of 13 ingredient groups from the complete IAbN. Child-of links are shown as arrows. Removed ingredient groups (fewer than ten ingredients) have a red background. **(b)** The Aggregate IAbN (b=10) for (a). Aggregate ingredient groups are shown as rounded-corner white rectangles, labeled with the numbers of ingredients, drugs and groups they summarize. (I = Ingredients; D = Drugs; G = Groups that were removed). Regular ingredient groups are white rectangles. They have the same I and D value in (a) and in (b).

Aggregate ingredient groups are not necessarily disjoint in terms of which removed ingredient groups they summarize. There may be two or more paths to different aggregate ingredient groups that satisfy the above condition. In Figure 2(a) Pipecolic Acids is a child-of Piperidines and Acids, Heterocyclic. On the path from Pipecolic Acids to Chemical Ingredients via Acids, Heterocyclic there are only removed ingredient groups. Thus, in the Aggregate IAbN of Figure 2(b), Pipecolic Acids is summarized by both Piperidines and Chemical Ingredients.

The child-of links between ingredient groups in the Aggregate IAbN are established according to the ingredient group hierarchy of the complete IAbN. If an aggregate ingredient group or a regular ingredient group is a child-of a removed ingredient group, then its child-of link(s) in the Aggregate IAbN go to the aggregate ingredient groups above the removed ingredient group. In other words, the child-of link is redirected to the aggregate ingredient group that summarizes the removed ingredient group.

IAbN Supporting the Discovery of New Candidates for Pharmacodynamic-based DDIs

One application of the IAbN is the discovery of new candidate drug-drug interactions (DDIs) not appearing in a DDI knowledgebase. Given is a set of DDI rules in the form (Drug$_1$, Drug$_2$, Clinical Consequence), such that Drug$_1$ and Drug$_2$ can be coded by NDF-RT drug ingredient concepts. In pharmacology, drugs with the same chemical ingredients tend to have similar DDIs [18]. By reviewing the DDIs associated with the drug ingredients in an IAbN ingredient group one may uncover new candidate pharmacodynamic-based DDIs.

To illustrate this process, we use an example from First Databank, Inc.'s DDI knowledgebase [19]. Consider the 18 drug ingredients in the Aggregate IAbN's Salicylates ingredient group (Figure 5(b)), 13 of which appear in FDB's DDI knowledgebase. The DDI interactions between ten of these salicylates and seven anticoagulant drugs are "Avoid concurrent use when possible" (AVD) and "Increases the effect of latter drug" (INL), for a total of 70 DDIs between these two groups. However, three extra Salicylates (balsalazide, mesalamine, and salsalate) have no DDIs with any anticoagulant in the FDB DDI knowledgebase. This raises doubt regarding the existence of DDIs between the seven anticoagulants and these three salicylates. Indeed, upon investigating the DDIs between the three extra salicylates and these seven anticoagulants in another source [20], we discovered DDIs between the three salicylates and three of the anticoagulants. The 70 old and nine new DDIs are described in Figures 3(a) and 3(b), respectively. The reason FDB did not include these new candidate DDIs in their knowledgebase is that in these cases the drug formulation has a low potential for interaction. Nevertheless, FDB staff (JKU) confirmed that this example

demonstrates the fact that summaries of NDF-RT have the potential for supporting the discovery of new candidate DDIs. Of course, pharmacological investigation is required for each potential DDI.

Results

We derived a (non-aggregate) IAbN for the February 2015 release of NDF-RT's Chemical Ingredients (CI) hierarchy which consists of 10,144 concepts. The complete IAbN consists of 859 ingredient groups which summarize 2,664 drug ingredients and 6,850 PP hierarchy drug concepts. We define the abstraction ratio of the IAbN to be the average number of drug ingredients per ingredient group. The abstraction ratio of the February 2015 IAbN is 3.1 (=2,664/859). There are 813 drug ingredient concepts summarized by more than one ingredient group and each such drug ingredient is summarized by an average of 2.25 ingredient groups. The average number of PP …
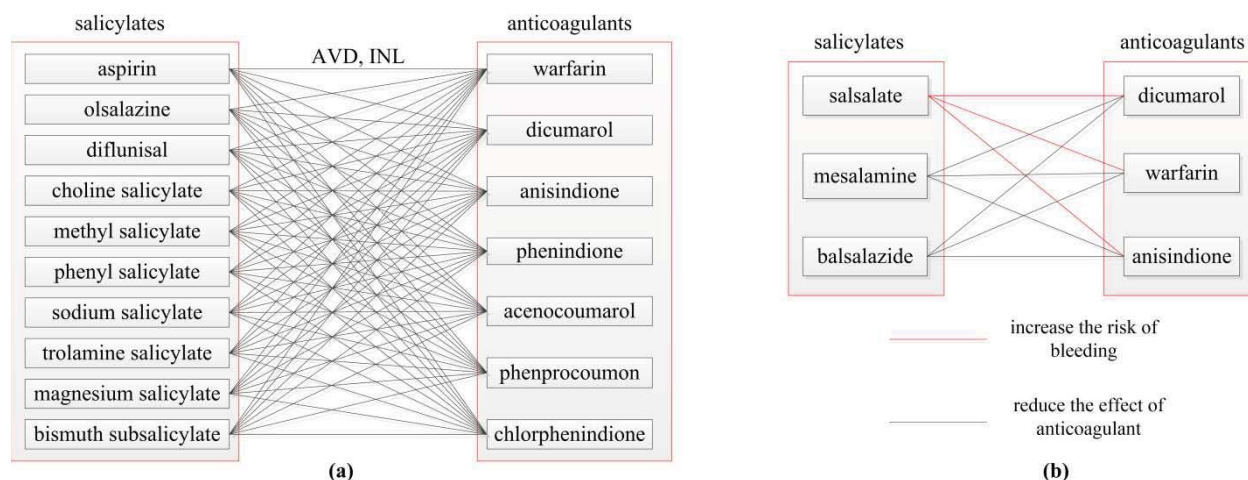

Figure 3(a). Illustration of 70 DDIs. There are 10x7=70 AVD and INL DDIs between the ten salicylates on the left and the seven anticoagulants on the right listed in FDB's DDI knowledgebase. **(b)** Nine new candidate DDIs not appearing in FDB's DDI knowledgebase. The DDIs of **(a)** combined with the aggregate ingredient group of salicylates in the IAbN of NDF-RT supported the discovery of nine DDIs that were confirmed by another source.

… drug concepts summarized by each ingredient group is 7.97. Figure 4 shows an excerpt of 149 of the IAbN's ingredient groups, as the complete IAbN is too large to fit on one page. By reviewing the ingredient groups of the IAbN one can see the major types of drug ingredients used in NDF-RT's drugs. For example, the Polymers group (Level 2: green) summarizes 26 ingredients and 81 drugs, Piperidines (Level 3: blue) summarizes 47 ingredients and 73 drugs, Tetracyclines summarizes 17 ingredients and 26 drugs, Ethanolamines summarizes 45 ingredients and 232 drugs, and Penicillins summarizes 34 ingredients and 64 drugs.

One deficiency of this complete IAbN is the relatively large number of "small" ingredient groups. For example, there are 312 (36%) ingredient groups that "summarize" only one drug ingredient. This is the reason why we developed the Aggregate IAbN methodology. The Aggregate IAbN can be used to combine these small "groups" into larger aggregate ingredient groups. We derived a few Aggregate IAbNs using bounds (b=) of 2, 6, 11, and 15, and investigated the structural properties of each resulting Aggregate IAbN (Table 1). The Aggregate IAbN in Figure 5(a), created using b=11, is composed of 30 regular ingredient groups (e.g., Ergotamines) and 88 aggregate ingredient groups (e.g., Saliycylates) which summarize 711 removed ingredient groups. Figure 5(b) illustrates how an aggregate ingredient group can be "dynamically expanded" with a software tool, so that its "summarized" removed ingredient groups can be recovered and viewed on demand.

IAbN-Based DDI Discovery Results

Table 2 shows the results of looking for new candidate DDIs using the Aggregate IAbN. Column 1 lists the interaction type. Column 2 represents the number of concepts of Group 1 (e.g., Salicylates) in NDF-RT. Column 3 represents the number of concepts in FDB among those in Column 2. Column 4 shows the number of concepts of Group 2 (e.g., Anticoagulants) interacting with Group 1 according to information in FDB's knowledge base. Column 5 describes the number of concepts of Group 1 having no interactions with Group 2 in FDB. Column 6 describes the number of concepts among those in Column 5 for which we have found drug interactions from public sources. Column 7 lists the number of drug interactions found in those public sources.

Chemical Ingredients (I:3) (D:21)

Level 2 (52 Ingredient Groups)

Group		
Amino Acids, Peptides, and Proteins (I:33) (D:106)	Polymers (I:26) (D:81)	Steroids, Fluorinated (I:19) (D:46)

- Amino Acids, Peptides, and Proteins (I:33) (D:106)
- Polymers (I:26) (D:81)
- Steroids, Fluorinated (I:19) (D:46)
- Nonsteroidal Anti-inflammatory Compounds (I:18) (D:26)
- Vitamin B Complex Compounds (I:18) (D:76)
- Estrenes (I:14) (D:31)
- Lipids (I:14) (D:107)
- Hexoses (I:13) (D:109)
- Norpregnenes (I:12) (D:40)
- Organic Chemicals (I:11) (D:43)
- Inorganic Chemicals (I:10) (D:329)
- Pregnadienes (I:9) (D:15)
- Androstanols (I:7) (D:8)
- Biological Products (I:6) (D:7)
- Unclassified Ingredients (I:6) (D:10)
- Biocompatible Materials (I:4) (D:5)
- Hexosamines (I:4) (D:37)
- Ketoses (I:4) (D:34)
- Polysaccharides (I:4) (D:4)
- Pregnenes (I:4) (D:6)
- Androstenes (I:3) (D:4)
- Cariostatic Agents (I:3) (D:37)
- Coenzymes (I:3) (D:4)
- Hydrolases (I:3) (D:22)
- Norpregnatrienes (I:3) (D:36)
- Tissue Extracts (I:3) (D:14)
- Antigens, Fungal (I:2) (D:5)
- Autacoids (I:2) (D:5)

Level 3 (191 Ingredient Groups)

- Piperidines (I:47) (D:73)
- Imidazoles (I:44) (D:73)
- Piperazines (I:36) (D:52)
- Pyridines (I:32) (D:227)
- Pregnenediones (I:22) (D:86)
- Pregnadienetriols (I:21) (D:51)
- Indoles (I:19) (D:32)
- Aminoglycosides (I:18) (D:53)
- Recombinant Proteins (I:17) (D:29)
- Plant Extracts (I:16) (D:111)
- Guanidines (I:15) (D:23)
- Pyrimidines (I:15) (D:36)
- Quinolines (I:15) (D:26)
- Benzimidazoles (I:14) (D:19)
- Estradiol Congeners (I:14) (D:31)
- Triazoles (I:14) (D:17)
- Organometallic Compounds (I:12) (D:50)
- Quinazolines (I:12) (D:17)
- Tetrazoles (I:12) (D:17)
- Aluminum Compounds (I:11) (D:43)
- Amino Acids, Diamino (I:10) (D:20)
- Amino Acids, Neutral (I:10) (D:19)
- Calcium Compounds (I:10) (D:68)
- Glucocorticoids (I:10) (D:10)
- Isoquinolines (I:10) (D:12)
- Potassium Compounds (I:10) (D:70)
- Sodium Compounds (I:10) (D:177)

Level 4 (266 Ingredient Groups)

- Phenothiazines (I:32) (D:83)
- Metals, Heavy (I:27) (D:303)
- Plant Oils (I:22) (D:69)
- Glucans (I:21) (D:57)
- Macrolides (I:20) (D:33)
- Transition Elements (I:19) (D:255)
- Tetracyclines (I:17) (D:26)
- Sugar Alcohols (I:16) (D:70)
- Aminopyridines (I:15) (D:65)
- Hydroxides (I:15) (D:83)
- Oligopeptides (I:15) (D:27)
- Acetates (I:14) (D:58)
- Dibenzazepines (I:14) (D:15)
- Ergotamines (I:14) (D:21)
- Ethylamines (I:14) (D:49)
- Thiazoles (I:14) (D:29)
- Benzodiazepines (I:12) (D:18)
- Fatty Alcohols (I:12) (D:32)
- Xanthines (I:12) (D:101)
- Anilides (I:11) (D:24)
- 17-Hydroxy-corticosteroids (I:11) (D:66)
- Dibenzocyclo-heptenes (I:11) (D:20)
- Ergolines (I:11) (D:24)
- Morpholines (I:11) (D:36)
- Naphthalenes (I:11) (D:17)
- Oxides (I:11) (D:113)
- Phenylacetates (I:11) (D:17)
- Thiophenes (I:11) (D:15)
- Aminoacridines (I:10) (D:22)

Level 5 (199 Ingredient Groups)

- Ethanolamines (I:45) (D:232)
- Phenethylamines (I:28) (D:81)
- Propanolamines (I:26) (D:159)
- Ethylene Glycols (I:24) (D:78)
- Sulfonamides (I:24) (D:39)
- Barbiturates (I:19) (D:116)
- Quaternary Ammonium Compounds (I:18) (D:38)
- Morphinans (I:17) (D:92)
- Benzhydryl Compounds (I:16) (D:56)
- Fluoroquinolones (I:14) (D:23)
- Chlorides (I:12) (D:200)
- Stilbenes (I:12) (D:14)
- Propylene Glycols (I:11) (D:43)
- Benzodiazepinones (I:10) (D:11)
- Pituitary Hormone-Releasing Hormones (I:10) (D:13)
- Anthraquinones (I:9) (D:22)

Level 6 (96 Ingredient Groups)

- Phenoxypropanolamines (I:20) (D:26)
- Catecholamines (I:19) (D:59)
- Sulfanilamides (I:16) (D:34)
- Morphine Derivatives (I:14) (D:171)
- Triiodobenzoic Acids (I:13) (D:72)
- Trimethyl Ammonium Compounds (I:13) (D:18)
- Amphetamines (I:12) (D:26)
- Antibodies, Monoclonal, Humanized (I:12) (D:14)
- Ethylenediamines (I:12) (D:31)
- Benzothiadiazines (I:11) (D:54)
- Sulfates (I:11) (D:112)
- Benzamides (I:9) (D:12)
- Tropanes (I:9) (D:41)
- Tryptamines (I:9) (D:13)
- Vinca Alkaloids (I:8) (D:11)
- Phenyl Ethers (I:7) (D:105)
- Phosphates (I:7) (D:39)
- Benzophenones (I:6) (D:37)

Level 7 (41 Ingredient Groups)

- Penicillins (I:34) (D:64)
- Cephalosporins (I:21) (D:39)
- para-Aminobenzoates (I:18) (D:104)
- Salicylates (I:13) (D:173)
- Retinoids (I:9) (D:42)
- Alkanesulfonates (I:6) (D:12)
- Scopolamine Derivatives (I:5) (D:27)
- Atropine Derivatives (I:4) (D:56)
- Alum Compounds (I:3) (D:11)
- Anisoles (I:3) (D:3)
- Calcium Phosphates (I:3) (D:5)
- Dihydroxy-phenylalanine (I:3) (D:8)
- Iothalamic Acid (I:3) (D:17)
- Ioxaglic Acid (I:3) (D:5)
- meta-Aminobenzoates (I:3) (D:4)
- Polyphosphates (I:3) (D:3)
- Amphenicols (I:2) (D:7)
- Arginine Vasopressin (I:2) (D:2)
- Bethanechol Compounds (I:2) (D:2)
- Bretylium Compounds (I:2) (D:3)

Level 8 (14 Ingredient Groups)

- Cephacetrile (I:14) (D:24)
- Cephamycins (I:6) (D:7)
- Aminosalicylic Acids (I:5) (D:7)
- Fenamates (I:4) (D:4)
- Phosphoramide Mustards (I:2) (D:7)
- Thienamycins (I:2) (D:3)
- Cephaloridine (I:1) (D:2)
- Clofibric Acid (I:1) (D:1)
- Mesylates (I:1) (D:1)
- p-Aminohippuric Acid (I:1) (D:1)

Figure 4. An excerpt of 149 (17%) ingredient groups from the February 2015 CI hierarchy's IAbN. The smaller ingredient groups have been hidden. Child-of links are hidden for readability and longer ingredient group names are truncated. The number of ingredients and drugs summarized by each ingredient group is shown in parentheses and prepended with I: and D:, respectively. Salicylates and Aminosalicylic acids, from Figure 1(c), are highlighted in yellow.
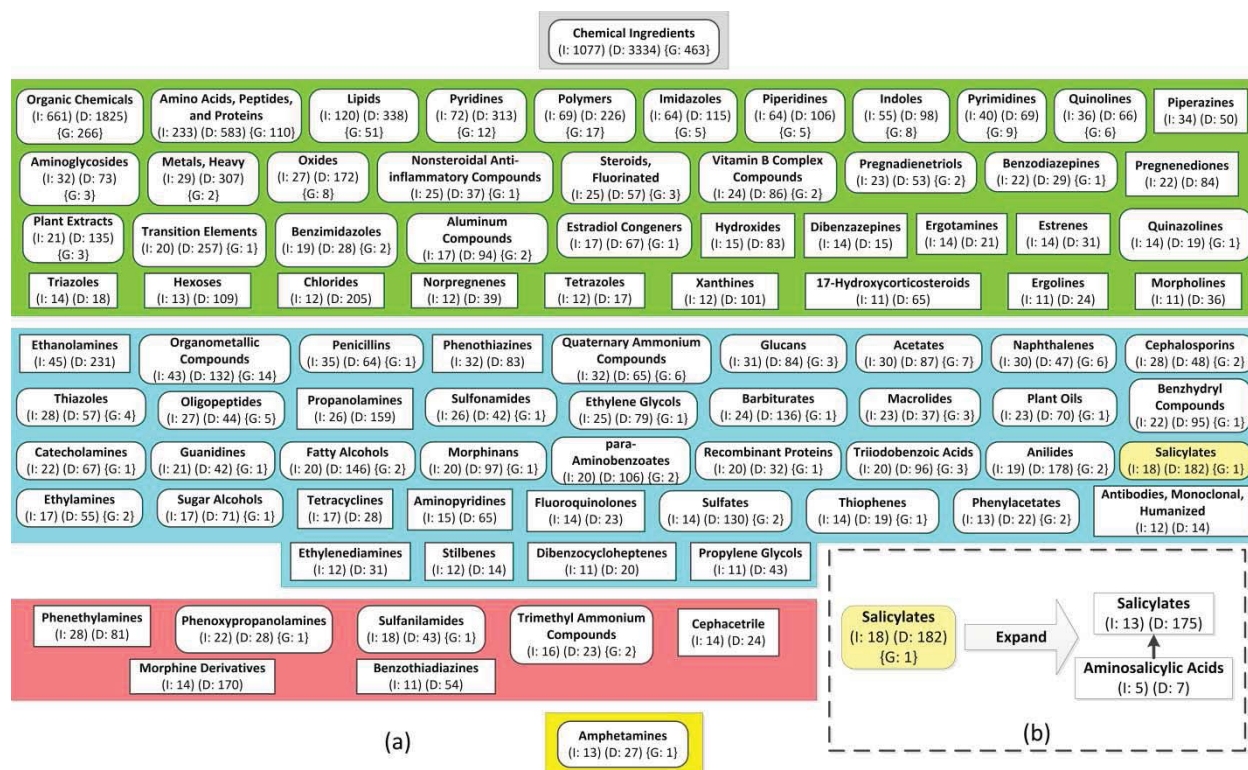
Figure 5. (a) The Aggregate IAbN (b=11) for CI. **(b)** The Salicylates aggregate group expanded.

Table 1. Structural metrics for Aggregate IAbNs created using various bounds.

Bound (b)	# Ingredient Groups (aggregate + regular)	# Aggregate Ingredient Groups (A)	# Removed Ingredient Groups (R)	Abstraction Ratio	R/A
No bound	859	0	0	3.10	-
2	547	138	312	4.87	2.26
6	184	102	675	14.5	6.61
11	88	58	771	30.3	13.29
15	45	37	814	59.2	22

Table 2. DDI discovery results

Column 1: Interaction Type	**Col 2:** #concepts of Group 1 in NDF-RT	**Col 3:** #concepts of Group 1 in FDB	**Col 4:** #concepts of Group 2 in FDB	**Col 5:** #Group 1 Concepts w/ no DDI in FDB	**C6:** #Col 5 concepts with new DDI candidates	**Col 7:** #new DDI candidates
Salicylates/Anticoagulants	18	13	7	3	3	9
Salicylates/Heparin	18	13	13	2	1	7
Salicylates/ Uricosurics	18	13	2	2	2	3
Salicylates/Antidiabetics, oral	18	13	9	3	3	24
Salicylates/ Valproic acid	18	13	5	4	2	3
Hydantoins/Selected anticoagulants	8	4	6	1	1	4
Hydantoins/ Isoniazid	8	4	1	1	1	1
Hydantoins/ Folic acid; Pyrimethamine	8	4	10	1	1	5
Hydantoins/Sulfonamides	8	4	4	1	1	4
Hydantoins/Cimetidine; Ranitidine	8	4	5	1	1	3

Table 2 lists five new candidate ingredient pairs (rows) for Salicylates. For each such pair Column 5 lists the number of salicylates from NDF-RT appearing in FDB but not having such a DDI. From those, Column 6 lists the number of salicylates that have DDIs with drugs of Group 2 in a public source and Column 7 lists the number of such DDIs. For example, row 1 reports the case described in Figure 3. However, the three extra salicylates e.g. mesalamine have no interactions with dicumarol, warfarin, and anisindione in FDB because their drug formulations produce localized concentrations in the gastrointestinal tract but not systemic blood levels where they will interact with anticoagulants.

In the last five rows of Table 2, we list ingredient pairs with hydantoin. Looking for hydantoins in NDF-RT that also appear in FDB but have no DDIs, a base formulation representation of the fosphenytoin ingredient in NDF-RT is found, for which public sources list DDIs. A review of FDB's knowledge base by a coauthor (JKU) found the DDIs associated with the salt formulation representation of this ingredient, fosphenytoin sodium, rather than with fosphenytoin. In this case, the candidate base ingredient does not exist in pharmaceutical formulations and thus there is no clinical data gap. However, there may be other instances, e.g., erthyromycin base with physical dosage forms that would be considered candidates for inclusion in DDIs. Hence, these DDI candidates are legitimate.

Discussion

The development of the IAbN (Ingredient Abstraction Network) represents an important first step in the summarization of NDF-RT's drug concepts according to their various classifications. The IAbN makes it possible to compactly visualize the major types of ingredients, and the drugs which contain them, as they exist in NDF-RT. One reason that the IAbN derivation methodology works well for the CI hierarchy is that the majority of drug ingredient concepts in NDF-RT are leaves (i.e., have no children) or are near the bottom (leaves) of the CI hierarchy.

The IAbN derivation approach is applicable beyond the CI hierarchy. For example, it is possible to apply the IAbN derivation methodology to the Mechanism of Action (MOA) hierarchy, summarizing NDF-RT's drugs according to their mechanisms of actions, rather than their chemical ingredients. In a future study, we will investigate the structural properties of IAbNs derived for NDF-RT's other classification hierarchies.

The Aggregate IAbN provides a secondary summarization mechanism that allows control over the granularity of summarization provided by an IAbN. The Aggregate IAbN lets a user see a very compact representation of a terminology hierarchy. Fine control is possible by choosing different values for b (the bound). The software tool (under development) for displaying IAbNs lets the user "re-expand" aggregate ingredient groups and lets her/him inspect previously hidden removed ingredient groups. In other words, the knowledge in the removed ingredient groups is not lost; it is just hidden and can be displayed by the tool upon demand. One issue to consider in future research is when one should use an Aggregate IAbN versus the complete IAbN. A usability study of the software tool and a study of the trade-offs between different abstraction ratios in Aggregate IAbNs will be performed.

In future research, we will investigate the use of alternate methods for controlling the amount of information of an IAbN that is displayed. This will include generating IAbNs that are created, e.g., by choosing a specific ingredient group and then viewing all of its ancestor and/or descendant ingredient groups. Additionally, we will investigate the creation of Aggregate IAbNs according to the number of PP drug concepts, rather than the number of ingredient concepts. This approach would provide an alternate secondary summary that highlights which ingredient groups summarize more PP drugs. Different users with various professional profiles may prefer different options according to their emphasis on ingredients or on drugs.

One significant difference between the IAbN and our previously developed abstraction networks is that not every concept in the CI hierarchy is summarized by an ingredient group. In our previous AbNs, every concept was always summarized by at least one AbN node [2]. However, since a CI hierarchy concept may be neither a classification ingredient nor a drug ingredient, it may be omitted from the IAbN, e.g., Ethyl biscoumacetate in Figure 1. In fact, the majority (6,621, 65.3%) of concepts in the CI hierarchy are not summarized by any ingredient group. This situation occurs for several reasons. The CI hierarchy was primarily imported from MeSH and many of the concepts from MeSH are too general and do not represent ingredients that could be used in drugs. Other ingredient concepts may not be relevant to the drugs in NDF-RT, as no drug includes them as an ingredient.

The goal of the IAbN is to summarize NDF-RT's drugs according to their ingredients. Thus, it is not necessary to summarize CI concepts that are not used as ingredients in drugs. However, it is conceivable that there may be applications which need access to these concepts. In a future study we will investigate ways of summarizing non-classification/non-drug ingredient concepts. One potential idea is to associate each such concept with its closest classification ingredient. Using such an approach, Ethyl biscoumacetate would be summarized by the 4-Hydroxycoumarins ingredient group.

Conclusions

In this paper, we introduced the Ingredient Abstraction Network (IAbN) to summarize the concepts in NDF-RT's Chemical Ingredients hierarchy. A parametric method for controlling the level of summarization in a secondary abstraction network, called Aggregate IAbN, was introduced. The IAbN was shown to support the discovery of new candidate drug-drug interactions in a DDI knowledgebase.

References

[1]   http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/, NDF-RT. [Cited 3/11/2015]

[2]   Halper, M., Gu, H., Perl, Y., et al. Abstraction Networks for Terminologies: Supporting Management of "Big Knowledge". Artificial Intelligence in Medicine. 2015:(accepted).

[3]   Wang, Y., Halper, M., Min, H., et al. Structural methodologies for auditing SNOMED. J Biomed Inform. 2007;40(5):561-81.

[4]   http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/VANDF/. [Cited 3/11/2015]

[5]   http://www.ncbi.nlm.nih.gov/mesh, MeSH. [Cited 3/11/2015]

[6]   Carter, J. S., Brown, S. H., Erlbaum, M. S., et al. Initializing the VA medication reference terminology using UMLS metathesaurus co-occurrences. Proc AMIA Symp. 2002:116-20.

[7]   Ochs, C., Geller, J., Perl, Y., et al. A Tribal Abstraction Network for SNOMED CT Hierarchies without Attribute Relationships. J Am Med Inform Assoc. 2014.

[8]   Wang, Y., Halper, M., Wei, D., et al. Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED. J Biomed Inform. 2012;45(1):15-29.

[9]   Min, H., Perl, Y., Chen, Y., et al. Auditing as part of the terminology design life cycle. J Am Med Inform Assoc. 2006;13(6):676-90.

[10]   Ochs, C., Perl, Y., Halper, M., et al. Gene Ontology Summarization to Support Visualization and Quality Assurance. BICoB 2015 (Accepted). 2015.

[11]   Stearns, M. Q., Price, C., Spackman, K. A., et al. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp. 2001:662-6.

[12]   Fragoso, G., de Coronado, S., Haber, M., et al. Overview and utilization of the NCI thesaurus. Comp Funct Genomics. 2004;5(8):648-54.

[13]   Ashburner, M., Ball, C. A., Blake, J. A., et al. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000;25(1):25-9.

[14]   Rosenbloom, S. T., Awad, J., Speroff, T., et al. Adequacy of representation of the National Drug File Reference Terminology Physiologic Effects reference hierarchy for commonly prescribed medications. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2003:569-78.

[15]   Carter, J. S., Brown, S. H., Bauer, B. A., et al. Categorical information in pharmaceutical terminologies. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2006:116-20.

[16]   Pathak, J., Murphy, S. P., Willaert, B. N., et al. Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2011;2011:1089-98.

[17]   Liu, S., Ma, W., Moore, R., et al. RxNorm: Prescription for Electronic Drug Information Exchange. IT Professional. 2005;7(5):17-23.

[18]   Brunton, L. L., Chabner, B. A., Knollmann, B. C. Chapter 3: Pharmacodynamics: Molecular Mechanisms of Drug Action. Goodman and Gilman's the pharmacological basis of therapeutics. 12 ed2011.

[19]   FDB MedKnowledge Clinical Modules: Drug-Drug Interaction Module: First Databank; 2014 Available from: http://www.fdbhealth.com/fdb-medknowledge-clinical-modules/drug-drug-interaction/.   [Cited 5/30/2014].

[20]   Drugs.com: Drug Interactions 2015 Available from: http://www.drugs.com/drug_interactions.php. [Cited 2/27/2015].