

A proposal for a drug information database and text templates for generating package inserts

Ryo Okuya¹
Masaomi Kimura²
Michiko Ohkura²
Fumito Tsuchiya³

¹Graduate School of Engineering and Science, ²Faculty of Engineering, Shibaura Institute of Technology, Tokyo, ³School of Pharmacy, International University of Health and Welfare, Tokyo, Japan

Abstract: To prevent prescription errors caused by information systems, a database to store complete and accurate drug information in a user-friendly format is needed. In previous studies, the primary method for obtaining data stored in a database is to extract drug information from package inserts by employing pattern matching or more sophisticated methods such as text mining. However, it is difficult to obtain a complete database because there is no strict rule concerning expressions used to describe drug information in package inserts. The authors' strategy was to first build a database and then automatically generate package inserts by embedding data in the database using templates. To create this database, the support of pharmaceutical companies to input accurate data is required. It is expected that this system will work, because these companies can earn merit for newly developed drugs to decrease the effort to create package inserts from scratch. This study designed the table schemata for the database and text templates to generate the package inserts. To handle the variety of drug-specific information in the package inserts, this information in drug composition descriptions was replaced with labels and the replacement descriptions utilizing cluster analysis were analyzed. To improve the method by which frequently repeated ingredient information and/or supplementary information are stored, the method was modified by introducing repeat tags in the templates to indicate repetition and improving the insertion of data into the database. The validity of this method was confirmed by inputting the drug information described in existing package inserts and checking that the method could regenerate the descriptions in the original package insert. In future research, the table schemata and text templates will be extended to regenerate other information in the package inserts.

Keywords: medical safety, drug information, package insert, drug database, cluster analysis

Introduction

According to a report published by The Japan Council¹ for Quality Health Care, 400 accidents and 180,000 incidents related to drugs occurred in 2010. In one case where a doctor prescribed 1800 mg of the total amount of a powder drug, the patient received 1800 mg of the active ingredient instead. To prevent such errors, it is necessary to check prescriptions and find errors.

It is expected that computerized systems in medical fields could prevent prescription errors by automatically checking prescriptions.² A computerized prescription order entry system is one of the computer systems that examine prescription data in a hospital. Because it manages the information of ordered drugs and patients, it should identify incorrect prescription information that can result in prescription errors. For example, if a doctor prescribes arotinolol hydrochloride tablets to a diabetic

Correspondence: Ryo Okuya
Shibaura Institute of Technology,
Toyosu 3-7-5, Koto-ku,
Tokyo 135-8548, Japan
Tel +81 3 5859 8507
Fax +81 3 5859 8507
Email mal1044@shibaura-it.ac.jp

patient, the computerized prescription order entry system should verify the indications for the drug and the symptoms of the patient. Because the indications for the drug are hypertension and essential tremor, the prescription is incorrect. In this situation, the system should warn the doctor, and he or she should consider other drugs. To create this system, it is necessary to create drug information databases that include principal drug data.

Most existing drug information databases are based on the descriptions in package inserts. Package inserts are the documents in which detailed drug information such as usage, dosage, and contraindications is described.^{3,4} The package inserts should be used as a primary source of drug data because they are officially published by pharmaceutical companies and authorized by authorities. However, it is difficult to extract drug data from the package inserts and create drug information databases on the basis of their information because there are no strict rules concerning the expressions used in package inserts. Even if package inserts contain the same information, they can be described in different formats such as tables, itemizations, and statements. In addition, because there is no regulation of medical vocabulary spelling, the phrasing in package inserts is not necessarily standardized. Therefore, it is difficult to create complete and accurate drug information databases by extracting drug information from package inserts.

Despite these issues, some studies have analyzed package inserts and the creation of drug information databases.⁵ Hamada et al proposed a method for efficiently retrieving drug information for package inserts by inserting tags into descriptions contained in the packets.⁶ However, this method is not a realistic approach if the number of drugs approved in Japan (>30,000) is considered because the tags have to be manually inserted into the descriptions. In a similar manner, Nabeta et al extracted active ingredient names from package inserts by pattern matching and obtained approximately 95% of the active ingredients.⁷ Nevertheless, even this method is not satisfactory because the drug information in all existing package inserts is needed. Because of the wide variety of expressions and formats in package inserts, it will be difficult to establish a systematic method for perfectly extracting drug data from package inserts.

To overcome these limitations, it was proposed to first build a database and then generate package inserts automatically by embedding data in the database into a suitable text template. The implemented database will not be affected by the wording of the package inserts, and the

database will help pharmaceutical companies because they will not need to create package inserts from scratch.

This study focused on the descriptions the drug composition, which is one of the principal components of package inserts, as a proof of concept. Database schema in a drug information database and text templates were created on the basis of the analysis of the drug composition descriptions. Drug composition descriptions were analyzed because they contain ingredient information, such as that for the active ingredient and any excipients, which can be utilized to prevent adverse events caused by improper drug combinations. Although it is crucial to avoid such drug combinations, it would be impossible to remember all such combinations because there are >30,000 drugs available in the Japanese market. After the drug ingredient information database is built, it will permit the implementation of functions to identify drug combinations regarded as contraindications.

In this paper, the descriptions were analyzed by cluster analysis, and the types of information and their description patterns in the package inserts are discussed. To manage the variety of drug-specific information in the package inserts, the information in the drug composition descriptions were replaced with labels. On the basis of the analysis, table schemata were designed, text templates were created to insert information in the database, and package inserts were generated.

Target data

The target data of this study are the descriptions in ethical drug package inserts published by the Pharmaceuticals and Medical Device Agency.⁸ A total of 11,547 package insert SGML files identified by the YJ code, provided by The Medical Information System Development Center,⁹ were downloaded.

It should be noted that the package inserts might include information on multiple drugs. In this study, package insert data corresponding to 14,639 drugs obtained by the corresponding division of the original files were used.

Methods

Drug information commonly appearing in package insert descriptions was identified by applying cluster analysis to the sets of words contained in the descriptions. On the basis of the results, table schemata in the drug information database were designed and text templates into which information in the database could be inserted were created.

In the following section, the analytical methods, selection standards of data attributes in table schemata, and methods of creating text templates are explained.

Analysis of description in package inserts

Generation of description patterns

To identify the common drug composition information included in package inserts, it is necessary to absorb the differences in expressions of similar information because of the difference originated in drugs.

In this study, such information was replaced with labels. The active ingredient names, excipient names, and contents, which were expressed by combinations of numbers and units, were replaced with labels because the Ministry of Health, Labor, and Welfare requires that this information is included in the composition section. In addition, the molecular formulas of the active ingredients and the property information such as dosage forms/colors in the descriptions were replaced. Table 1 shows a list of the word types that were replaced and their replacement labels.

After the replacement, unnecessary characters and delimiters (eg, spaces, punctuation, and carriage returns) in these descriptions were deleted.

Hereafter, the resultant descriptions are called “patterns” (Figure 1).

Cluster analysis

To group similar patterns, cluster analysis was applied to the generated patterns.

Cluster analysis methods were used to classify similar data into groups called clusters. Their algorithms are categorized

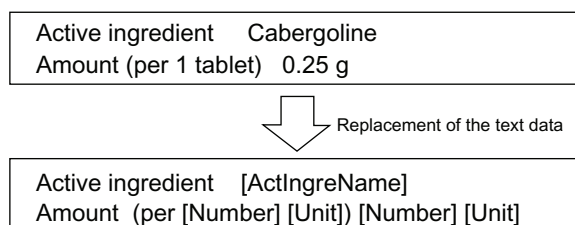


Figure 1 An example of the generation of description patterns.

into two types: partitioning–optimization and hierarchical types. Partitioning–optimization cluster analysis classifies similar data into exclusive clusters, whereas hierarchical cluster analysis aggregates similar data in sequence according to the distance between them and describes the aggregation process as a tree diagram called a dendrogram. The height of a dendrogram corresponds to the distance within which data are clustered. For example, if the dendrogram is cut at an arbitrary height, clusters within which the distances of any data are less than the total height are obtained.

In this study, cluster analysis was performed in two steps. The first step was based on the K-means algorithm, one of the partitioning–optimization methods, to identify the same patterns as preclusters. The second step was hierarchical cluster analysis, which was used to identify “clusters” of preclusters using a dendrogram.

The distance between patterns must be defined when employing hierarchical cluster analysis. This study focused on nouns, which are the most numerous component of the patterns and should be used to classify the patterns, because the same words tend to appear in similar patterns. Vectors representing the patterns were created, the elements of which denote the existence of nouns or labels. If a noun or one of the labels exists in the pattern, the correspondent element is one; otherwise, it is zero.

To define the vector distances, a problem known as the “curse of dimensionality,” which occurs because of the high dimensionality of the vectors, must be managed. In this study, the vector dimensions were decreased by applying singular value decomposition to the matrix, each row of which is the previously defined vector.

The execution of cluster analysis requires distances between data (patterns in this study). The distance (more strictly, pseudodistance in this study) was calculated on the basis of the cosine value of the angle between a pair of vectors. Although the cosine value is suitable for measuring the similarity of the vectors, it lacks the property of a distance, as the similarity of the vectors increases as the cosine

Table 1 The types of words that were replaced and their corresponding labels

Drug information	Label
Brand name	[BrandName]
Active ingredient name	[ActIngreName]
Excipient name	[ExcipientName]
Other name of active ingredient	[ActIngreOtherName]
Abbreviated name of active ingredient	[ActIngreShortName]
Number	[Number]
Unit	[Unit]
Chemical formula	[Formula]
Peculiar part of drug (eg, support medium)	[DrugPart]
Color	[Color]
Dosage form (eg, injection)	[Form]
Smell	[Smell]
Taste	[Taste]
Usage of excipient	[Usage]

value increases. Thus, the cosine value was converted into a distance by subtracting it from one and applied hierarchical cluster analysis on the basis of this distance.

Database

The results of the analysis identify what drug information should be stored in the database. The results illustrate the common drug information in many package inserts, and on the basis of these results, data attributes for the tables were defined. It should be noted that not all of the obtained information can be entered as data attributes. Information must be selected from the viewpoint of utilization in a computerized system, such as the automatic checking system of prescription orders.

The data attributes were combined into tables and submitted to database normalization.

Text template

The strategy for generating a package insert was to embed the data in the database into a text template. The description patterns were regarded as text templates because they are based on original and typical package inserts. Labels in the patterns were regarded as positions to assign drug data in the database.

Multiple templates can be obtained from all description patterns. To decrease the number of templates, the results of the cluster analysis were utilized. Typical description patterns that appeared in the largest number of package inserts in a cluster were regarded as text templates.

One might find it reasonable to create templates from a random subset of package insert descriptions. However, the results of such a strategy potentially lack templates to insert characteristic drug information in infrequently appearing description patterns. Consequently, this approach was not adopted.

To generate an appropriate package insert, it is necessary to determine the proper text template according to the drug information to be described in the package insert. A flag was assigned to drug information and the patterns of the flags were assigned to the templates.

Results

Analysis of description in package inserts

Generation of description patterns

Concerning the replacement results for keywords in the descriptions, 3046 active ingredient information patterns, 1591 excipient information patterns, and 187 biological ingredient information patterns were generated. Consulting the

active ingredient patterns, the patterns of blood preparations and radiopharmaceuticals contained many words that were not replaced with labels (Figure 2). This suggests that such words lead to unsatisfactory pattern categorization; the patterns that carry similar information can be categorized into different clusters as a consequence. Because of this, the patterns of blood preparations and radiopharmaceuticals derived from other patterns were analyzed. The same issue was also observed for the patterns of Chinese herbal drugs, which usually contain ingredients with unknown biochemical functions. These patterns were called peculiar patterns. From this viewpoint, the patterns were divided into 2685 active ingredient patterns and 379 peculiar patterns.

Cluster analysis

Figure 3 shows the dendrogram that represents the clustering result of the active ingredient information patterns. The red line in Figure 3 shows the height of the dendrogram at which the patterns were divided into clusters. The height was selected to obtain identical patterns in each cluster. As a result, ten clusters were obtained.

Table 2 shows examples of the patterns included in the largest cluster (Cluster 2) in Figure 3. It was found that most patterns consisted of an active ingredient name, its content, and a dosage unit expressed as “per [Number] [Unit].” Moreover, patterns in which the expression “[Number] [Unit] of [ActIngreName]” appeared repeatedly as well as the same expressions written in parentheses were also found. The former described the content of the active ingredients, and the latter described compounds including the active ingredient such as salts and hydrates.

Table 3 shows typical description patterns in the clusters excluding the largest cluster. The patterns in some clusters included information about the raw materials of active ingredients such as “flour” and supplementary information of the active ingredient such as “recombinant” after an expression of the active ingredient name. It can be seen that patterns in different clusters include the terms “ion concentration” and “calorie.” A pattern in another cluster contains the expression “not less than” described with the content of an active ingredient. This describes the lower bound of the content of active ingredients in the drugs. Moreover, some patterns

<p>Active ingredient (per [Number] [Unit]) [ActIngreName] alpha (activated) (recombinant) [Number] [Unit] [Number] [Unit] of active ingredients correspond to 50 KIU</p>
--

Figure 2 An example of blood preparation description patterns.

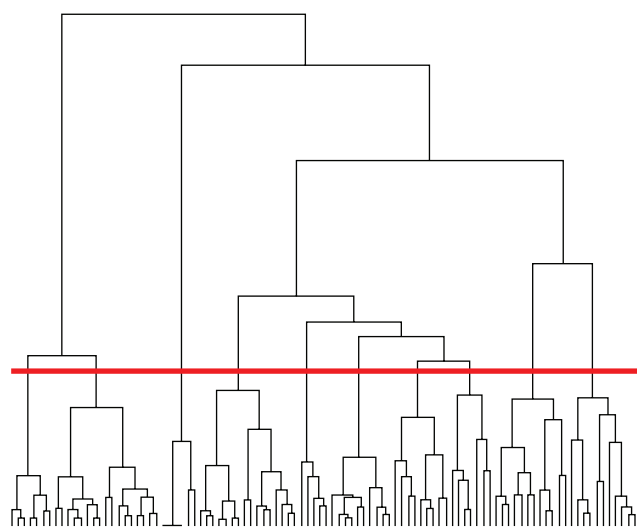


Figure 3 A dendrogram of active ingredient description patterns.

included information about production methods such as “dialysis” and “aseptic manipulation” and about the usage of a drug such as “dissolve before use.”

Concerning peculiar patterns, an additional ten clusters were found.

Consulting the patterns for Chinese herbal drugs, it was found that the patterns contained information about herbal drug content as well as multiple herb names and their contents. Table 4 shows examples of patterns in clusters excluding the clusters that included the patterns of Chinese herbal drugs. Although most patterns are the same as those shown in Tables 2 and 3, the patterns in some clusters describe the radioisotope information of an active ingredient such as “^{99m}Tc.”

Figure 4 shows the dendrogram that represents the clustering result of excipient patterns. The red line in

Table 2 Examples of typical active ingredient description patterns in Cluster 2

Description patterns
Active ingredient (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]
Ingredient amount (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]
Active ingredient (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]
[Number] [Unit] of [ActIngreName] [Number] [Unit] of [ActIngreName]
Ingredient amount (per [Number] [Unit])
[Number] [Unit] of [ActIngreName] ([Number] [Unit] of [ActIngreName])
Volume [Number] [Unit] Ingredient amount (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]

Table 3 Examples of typical active ingredient description patterns in clusters excluding Cluster 2

Cluster number	Description patterns
1	Brand name [BrandName] Active ingredient and its amount (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]
3	This drug which contains a specific allergen obtained by dialysis from the extraction ingredient from material (flour) is aseptic physiological saline This drug regulated by aseptic manipulation is [Number] [Unit] [Form] for the material weight ([Number]: [Number])
4	The following ingredients are contained per [Number] [Unit] Active ingredient [Number] [Unit] of [ActIngreName]
5	Active ingredient (per [Number] [Unit]) [Number] [Unit] of [ActIngreName] (recombinant)
6	This product contains the following ingredients per [Number] [Unit] Active ingredient and its amount [Number] [Unit] of [ActIngreName]
7	[BrandName] which contains [ActIngreName] [Number] [Unit] per [Number] [Unit] is administrated to dissolve before use Active ingredient [Number] [Unit] of [ActIngreName]
8	[BrandName] which contains [ActIngreName] [Number] [Unit] per [Number] [Unit] is [Form] of [Color] <Electrolyte concentration> (theoretical value) [Number] [Unit] of [Formula]+
9	[ActIngreName] [Number] [Unit] Calorie [Number] [Unit]
10	This product contains not less than [Number] [Unit] of [ActIngreName] ([Formula])

Figure 4 shows the height of the dendrogram at which the patterns were divided into clusters. The height was selected to obtain identical patterns in each cluster. As a result, seven clusters were obtained.

Table 5 shows the typical patterns in each cluster. It was found that excipient names repeatedly appeared in patterns in all clusters. In addition, clusters for which their patterns included information about excipient content and usage and clusters for which their patterns did not explicitly include excipient names but did include their number were found.

Concerning the biological ingredient patterns, seven clusters were obtained from the dendrogram. The patterns

Table 4 Examples of typical peculiar description patterns

Description patterns
Active ingredient (per [Number] [Unit]) [Number] [Unit] of [ActIngreName]
The following ingredients are contained per [Number] [Unit] Active ingredient [Number] [Unit] of [ActIngreName]
[Number] [Unit] of [ActIngreName] ([Formula]) solution (assay date) per [Number][Unit]

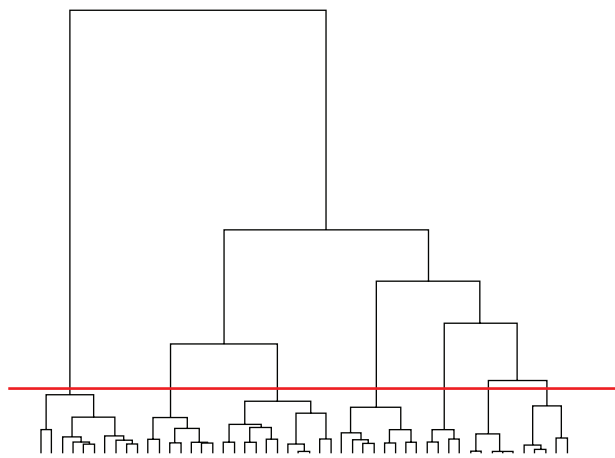


Figure 4 A dendrogram of excipient description patterns.

appearing in all clusters included expressions of information designated in the official guideline by Ministry of Health, Labor, and Welfare. The characteristic patterns included information about ingredients used in manufacturing such as “cell lines of Chinese hamster ovary were used to this product at a cultured process.”

Database

The names of active ingredients, their contents, and the dosage unit are essential data attributes in a drug information database. This information is useful for preventing drug overdoses. Moreover, as an active ingredient and its content can appear together in pairs, the table schemata must store such pairs. Regarding the content of the active ingredient, the result suggested that it could be described as a range; specifically, the upper/lower limit values of the content

Table 5 Examples of typical excipient description patterns
Description patterns
Excipient [Number] [Unit] of [ExcipientName] [Number] [Unit] of [ExcipientName]
[Number] [Unit] of [ExcipientName] Provided solution [Number] [Unit] of [ExcipientName]
Excipient [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName]
[DrugParts] [ExcipientName] [ExcipientName] [ExcipientName]
Excipient (per [Number] [Unit]) [Number] [Unit] of [ExcipientName]
Excipient [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName]
[ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName]
[ExcipientName] and [Number] other ingredients
Excipient [Number] [Unit] of [ExcipientName]
Excipient [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName]
This product contains [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName] [ExcipientName] as excipients

should be stored in the database. Although the salt/hydrate and ion concentration data are optional, they must be included as attributes in the schemata because they are important for representing the substance closely tied with the active ingredient name information. In a similar manner, other information such as Chinese herbal drug and radioisotope information should be included as attributes.

Conversely, the usage of a drug was not defined as an attribute because this information should be described in the section “dosage and administration.” The attribute of production method information was excluded from the attributes because it is also irrelevant to the composition information of an ingredient.

Regarding information on the excipient, the name and content of an excipient are important attributes. Because the pairs of names and contents also repeatedly appeared in most patterns, table schemata were designed to store such repeated pairs.

From the results of cluster analysis of biological ingredient patterns, it was elucidated that some patterns described information on ingredients used in manufacturing in addition to the biological ingredient. Therefore, it is necessary to design data attributes to store the ingredient name, its content, and the name of the process in which an ingredient is used.

Then, codes were added to identify drugs and ingredients to the attributes of tables and submitted them to applied database normalization. There are two types of attributes: attributes related to each ingredient in the drug and those related to the ingredients as a whole. These types of attributes were stored in separate tables. In addition, tables were designed to store other information, such as those of ions and Chinese herbal drugs.

On the basis of the database normalization, table schemata were designed in consideration of the relationships of the tables. Figure 5 shows a portion of the schemata.

Text template

Text templates were created on the basis of the description patterns shown in Tables 2–5. However, this method does not allow description patterns in which repeated data can be inserted. In addition, it is difficult to cover all of the appearance patterns of optional data. To solve this problem, repeat tags were introduced in the templates to indicate repetition. When a package insert was generated, the repeated data was embedded into the part surrounded by the repeat tag with an arbitrary number of repetitions.

The result indicated that no description pattern could encompass all nutrient drug data. Consequently, the parts of

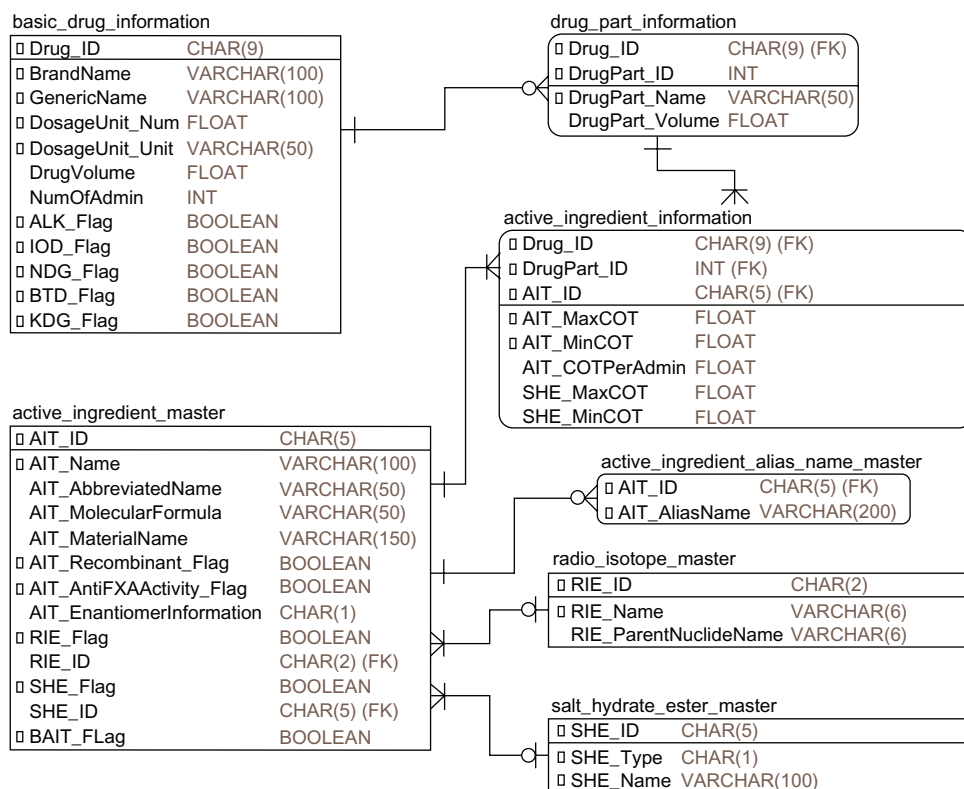


Figure 5 Entity relationship diagram based on the proposed data schemata (main part).

the templates were arranged corresponding to nutrient drug components and combined to generate package inserts.

Table 6 shows examples of the templates. The Attr attribute in the repeat tags indicates the types of target information. As shown in Table 6, many templates can include active ingredient information and primary drug information such as dosage units.

Figure 6 shows an example of the used template and a description including information. “Ait” indicates that the repetition should be the number of active ingredients. In this example, active ingredients and their contents were repeated three times.

Experiment

To evaluate the validity of the proposed method, drug information described in existing package inserts was inputted into the proposed database and it was verified that the method could regenerate the descriptions in the original package inserts. Two-hundred randomly selected package inserts were used.

From the experimental results, it was confirmed that the method succeeded in storing drug information for 178 package inserts (89.0%). The data to be stored that did not exist but could be derived from other data in package inserts

were regarded as successes. For example, if active ingredient information was not available but salt/hydrate information was available, this case was regarded as a success because the amount of the active ingredient could be derived from the salt/hydrate information.

Package inserts were regenerated for the data successfully stored in the database. Comparing regenerated

Table 6 Examples of text templates

Templates

Ingredient content (per {Dosage Unit})
<repeat attr = "ait"> {Active ingredient content} of {Active ingredient name} </repeat>

Volume {Drug Volume}

Ingredient content (per {Dosage Unit})
<repeat attr = "ait"> {Active ingredient Content} of {Active ingredient name} </repeat>

Active ingredient (per {Dosage Unit} in {Drug Volume})
<repeat attr = "ait"> {Active ingredient Content} of {Active ingredient name} </repeat>

<repeat attr = "ion"> Ion concentration {Ion concentration} of {Ion name} </repeat>

Excipient
<repeat attr = "exc"> {Excipient name} </repeat>

Excipient (per {Dosage Unit})
<repeat attr = "exc"> {Excipient content} of {Excipient name} </repeat>

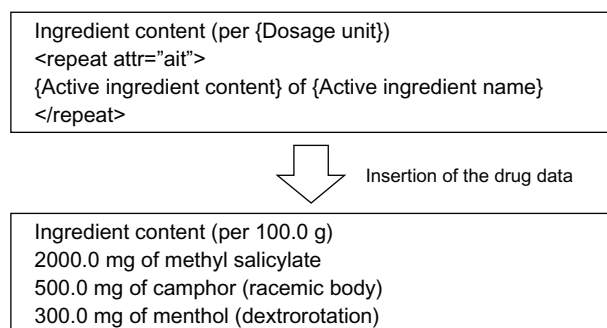


Figure 6 An example description including information and the used template.

and original package inserts, it was confirmed that the packets were properly regenerated. It should be noted that in 14 cases, the expressions in the regenerated package inserts were not identical to those in the original inserts. This issue originated from the difference in the units of drug information, for example, the use of mass concentration (%) in the original package inserts and “mg” in the database. The descriptions in these cases can be considered to regenerate original drug information because the drug volume is described in the package inserts and the mass can be derived.

Impact on medical safety

The drug information databases that were proposed will enable the creation of a computerized system to prevent prescription errors. Utilizing this database, the system can obtain active ingredient information about the drugs.

Utilizing the obtained active ingredient information, the system determines whether the ordered drugs are properly selected according to patient symptoms. On the basis of the active ingredient codes, the system obtains indications corresponding to the active ingredient from an indication information table (Table 7). As a result, the system reports “hypertension” and “essential tremor” as the indications for arotinolol hydrochloride and “migraine” and “orthostatic hypotension” as the indications for dihydroergotamine mesylate. Upon verifying the indications and patient

Table 7 Sample indication information table

AIT_ID	Indication
00001	Insomnia
00001	Focal seizure
00002	Hypertension
00002	Essential tremor
00003	Migraine
00003	Orthostatic hypotension
00004	Type 2 diabetes

symptom information, the system reveals that hypertension is one of the indications for arotinolol hydrochloride, but the indications for dihydroergotamine mesylate are not matched. Therefore, because the system finds that dihydroergotamine mesylate tablets were incorrectly ordered, it cancels one prescription order and alerts the doctor.

Conclusion

In this study, to create a computerized system to prevent prescription errors, a new strategy for building a drug information database was proposed.

The drug composition descriptions in package inserts were analyzed. On the basis of the results, table schemata for the drug information database and text templates to which drug-specific data in the database was embedded were proposed.

Consequently, the data attributes (eg, the salt/hydrate of an active ingredient), the supplemental information of an active ingredient, and the usage of an excipient were found, in addition to the information required by the package insert description guideline in Japan. Drug information that appears repeatedly in drug descriptions was also found. To reflect this fact, the data attributes of table schemata were arranged to store such drug information. Text templates were created, into which drug data were inserted, and repeat tags were generated to reflect the repetition of the corresponding drug information.

From the experimental results, it was confirmed that most of descriptions in package inserts could be regenerated from the proposed database and text templates.

In future research, the table schemata and text templates will be extended to regenerate descriptions for other information in package inserts.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Tokyo: Japan Council for Quality Health Care. *Project to Collect Medical Near-miss/ Adverse Event Information 2010 Annual Report*. Tokyo: Japan Council for Quality Health Care; Aug 30, 2011. Available from http://www.med-safe.jp/pdf/year_report_english_2010.pdf. Accessed January 11, 2013.
2. Sugiyama T, Niwa T, Takagi N, Goto C, Katagiri Y. Investigating the usefulness of a prescription checking system in risk management. *Japanese Journal of Pharmaceutical Health Care and Sciences*. 2003; 29(1):73–76. Japanese.
3. Ministry of Health, Labor, and Welfare. *A Guideline for Descriptions in Ethical Drug Package Inserts*. Tokyo: Ministry of Health, Labor, and Welfare; 1997. Japanese.
4. Ministry of Health, Labor, and Welfare. *A Guideline for Descriptions in Biological Drug Package Inserts*. Tokyo: Ministry of Health, Labor, and Welfare; 2003. Japanese.

5. Togashi H, Kuribara M, Orii T. Analysis for contents data of package leaflet in medicinal supplies. *Japan Journal of Medical Informatics*. 2006;26(2):129–134. Japanese.
6. Hamada M, Hirota M, Kurata K, Dobashi A. Creation of drug information database based on pharmaceutical markup language (PML). *Japanese Journal of Pharmaceutical Health Care and Sciences*. 2007; 33(6):502–509. Japanese.
7. Nabeta K, Kimura M, Ohkura M, Tsuchiya F. A proposal of a method to extract active ingredient names from package inserts. *Lecture Notes in Computer Science*. 2009;5618:576–585. Japanese.
8. Menu of package insert information [webpage on the Internet]. Tokyo: Pharmaceuticals and Medical Devices Agency; [cited May 26, 2011]. Available from: <http://www.pmda.go.jp/>. Accessed January 11, 2013. Japanese.
9. HOT code master [webpage on the Internet]. Tokyo: The Medical Information System Development Center; [cited March 31, 2011]. Available from: <http://www.medis.or.jp/>. Accessed January 11, 2013. Japanese.

Drug, Healthcare and Patient Safety

Publish your work in this journal

Drug, Healthcare and Patient Safety is an international, peer-reviewed open-access journal exploring patient safety issues in the healthcare continuum from diagnostic and screening interventions through to treatment, drug therapy and surgery. The journal is characterized by the rapid reporting of reviews, original research, clinical, epidemiological and

Submit your manuscript here: <http://www.dovepress.com/drug-healthcare-and-patient-safety-journal>

post-marketing surveillance studies, risk management, health literacy and educational programs across all areas of healthcare delivery. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress