공학석사 학위논문

iDrug: Text mining driven international adverse drug-drug interactions

iDrug: 텍스트마이닝 기반 약물 상호작용 분석

2016 년 2 월

서울대학교 대학원

전기·정보공학부

김 세 정

국문초록

눈부신 의학 분석 기술의 발전으로 수많은 약물-약물 상호관계가 규명되었다. 이러한 상호관계 중 약물-약물 금기관계는 함께 복용 시 부작용의 강도가 심한 것으로 Intensice Care Unit (ICU)와 같은 응급상황에서 의료 전문가들에게 우선적으로 주어져야 하는 정보이다. 기존의 약물-약물 금기관계를 제공하는 drug compendia는 서로 다른 compendia간의 약물-약물 금기관계 리스트 불일치가 문제되고 있으며, 현재 이러한 약물-약물 금기관계를 규명하고, 구조화하는 연구가 미흡한 상황이다.

우리는 텍스트마이닝을 적용하여 6개국 (미국, 캐나다, 영국, 프랑스, 스위스, 일본) Prescribing Information(PI)로부터 약물-약물 상호관계를 추출하고, 약물-약물 금기관계를 제공하는 국가별 약물-약물 상호관계 심각도 데이터베이스를 구축하였다. 또한, 추출한 정보를 바탕으로 국가별 약물-약물 상호관계 그래프 특성을 파악하고, 국가 간 약물-약물 금기관계 차이를 보였다. 본 연구를 통해 6개국 국가 간 약물-약물 금기관계 리스트가 0.6 이하의 overlap을 가짐을 보였으며, intraclass coefficient correlation이

0.27로 낮은 상관관계를 가짐을 보였다. 본 연구는 6개 국가의 국가 규제 기관에서 승인하고 제공하는 PI를 text mining을 통해 약물-약물 상호관계를 추출하고 심각도 데이터베이스를 구축하여 그 결과를 비교하는 첫 시도이다. 또한, 국가 규제 기관에서 승인한 문서를 바탕으로 국가별 약물-약물 상호관계 심각도를 구축하였기 때문에 의료관계자들에게 가이드라인 역할을 할 수 있다는 데 의의가 있다.

주요어 : 약물-약물 금기관계, 텍스트 마이닝, 그래프 비교

학 번 : 2014-21735

Table of Contents

List of Tables

LIST OF FIGURES

Ⅰ. Introduction

Drug-Drug interactions (DDIs) occur when multiple drugs are being administrated. DDIs can be accompanied by various adverse drug reactions, such as fever and abdominal pain, and often cause severe consequences in patients, from hospitalization to death. Therefore, it is important to discover adverse DDIs with their severity and construct a DDI severity database based on this information [1]. The severity of DDIs should be considered by clinicians when prescribing drugs; for example, a DDI that causes difficulty in breathing should have more severity than a DDI that causes itching. Moreover, 49~69% of DDI alerts that occur in the intensive care unit (ICU) are overridden by clinicians because of alert fatigue, a phenomenon where people become desensitized to alerts when they are frequently exposed to a large number of alarms [2 - 3]. Only 11% of DDI alerts are considered useful. Thus, it is important to reduce the number of alerts by discovering and prioritizing severe adverse DDIs. In this study, we created a DDI severity database to effectively prevent the occurrence of adverse drug reactions.

The main sources of DDI information can be categorized into two

types: (1) commercial databases and (2) public databases. Commercial databases, such as Micromedex Drug-REAX [4], Lexi-comp [5], the Veterans Affairs (VA) system [6], and Drug Interactions: Analysis and Management (DIAM) [7], are created manually by various healthcare providers and have been widely used in the field. They provide a DDI severity rating system, which is partly embedded in a clinical decision support system (CDSS). However, discrepancies in the levels of agreement in DDI severity ratings between compendia have recently become problematic. A study in which researchers investigated the agreement of adverse DDIs in three compendia (VA, Micromedex, and DIAM) showed that only 13.7% of interactions were listed in all three sources [8]. Only 5% of DDIs described in the FDA black box warning of 11 drugs also were rated as contraindicated in Lexci-comp, Micromedex, and Facts & Comparisons [9]. Public databases, such as Drugbank [10] and Kyoto Encyclopedia of Genes and Genomes (KEGG) Drug [11], are freely available online resources providing comprehensive drug information (i.e., chemical, pharmacological, and pharmaceutical information). Although DDI information provided by the Drugbank has been widely used as a golden-standard, it does not classify the severity of DDIs such as

adverse DDIs. KEGG Drug is an integrated resource providing information about approved drugs in the United States, Europe, and Japan. It divides DDI severity into two classes, namely contraindications and precautions, based on the text mining results of the prescribing information provided by the Japan Pharmaceutical Information Center (JAPIC) [12]. Nevertheless, it is limited by containing only adverse DDIs mined from documents related to drugs marketed in Japan. Previous studies show that matches of adverse DDIs between different compendia are quite low, and it is believed that problems exist in referencing only one resource such as the documents relating to one country or a single drug compendium.

In this study, we applied text-mining techniques to extract drug-drug interaction data from prescribing information in six countries—the United States, the United Kingdom, Canada, France, Switzerland, and Japan—and constructed an integrated DDI severity database that stores two levels of DDI severity, contraindications and precautions, for the six countries. Prescribing information (PI) is officially approved by drug agencies such as the European Medicines Agency (EMA) and US Food and Drug Administration (FDA) and provides authoritative information for the safe prescription and

indication of a drug [13]. License holders are required by law to provide PI for each drug marketed. Since our data were populated from the documents approved by national regulatory authorities, we believe that iDrug has high reliability and that it represents the features of DDIs for each nation. Based on the information we extracted, we analyzed the characteristics of DDI graphs for each nation. We also computed the pairwise differences of adverse DDIs between nations to understand the inconsistency level of the DDIs of each country compared to other countries.

To the best of our knowledge, our work is the first attempt to automatically extract PI from the national regulatory authorities of the six countries using text-mining techniques. As our data were populated from the documents approved by national regulatory authorities, we believe our database can serve as a guideline for medical experts.

Ⅱ. Materials and Methods

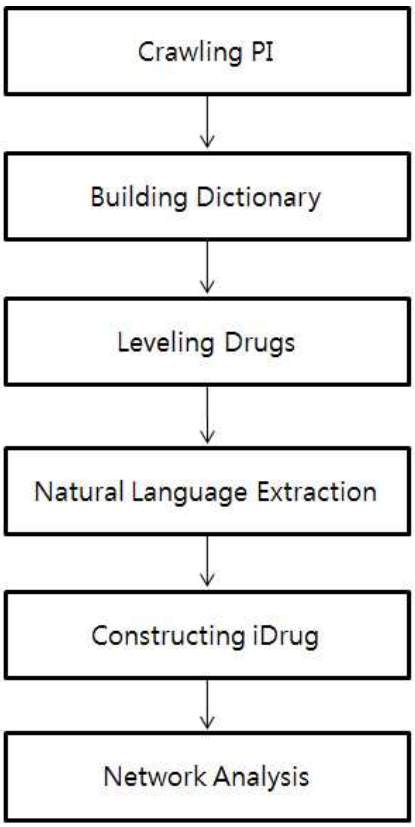

Figure 1 An overview of our approach

Figure 1 shows the overall flow of the proposed iDrug approach, which consists of five main steps: (1) crawling PI, (2) building a dictionary of drug substances, (3) hierarchical leveling of drugs, (4) text-mining driven drug-drug interactions extraction, and (5) adverse drug-drug interaction network analysis with iDrug

2.1 Crawling Prescribing Information

As shown in Table 1, we crawled PI from the representative drug compendia of five countries (the United States, the United Kingdom, Canada, France, and Switzerland) and excluded Japan. Drug compendia such as Health Canada, eMC, Sante, and Documed allow users to access data and reference documents on marketed drugs or drugs traded during the last three years in each country; the information is supplied in PDF or html format. Dailymed provides structured product labeling (SPL), which is an electronic format for PI. In the case of Japan, we used KEGG Drug DDI severity for the network analysis and integrated it in iDrug. As mentioned above, KEGG Drug extracts two levels of DDI severity (contraindications and precautions) by applying text mining to PI provided by the JAPIC. At the time, in July 2015, Dailymed, eMC, Health Canada, Documed, and Sante had provided approved PI for 42,731; 5,240; 7,435; 7,304; and 8,814 drugs.

Table 1 Data sources for crawling prescribing information and extracting drug-drug interaction from six countries

Country	Data source	Doc type	Language
United States	Dailymed (http://dailymed.nlm.nih.gov/)	XML	English
Canada	Health Canada (http://webprod5.hc-sc.gc.ca/dpd-bdpp/)	PDF	English
United Kingdom	eMC (http://www.medicines.org.uk/emc/)	HTML	English
France	Sante (http://base-donnees-publique.medicaments.gouv.fr/)	HTML + PDF	French
Switzerland	Documed (http://compendium.ch/)	HTML	French
Japan	KEGG Drug (http://www.genome.jp/kegg/drug/)		

2.2 Drug Dictionary and Hierarchical Leveling

We have built a drug substance dictionary with international nonproprietary names (INNs) as the main references. INNs, developed and maintained by the World Health Organization (WHO), are the unique, globally recognized names of pharmaceutical substances or active pharmaceutical ingredients [14]. This naming system was devised to avoid prescription errors due to the communication of drug names among clinicians when brand names of two drugs are different but active ingredients are identical. As of June 2015, the WHO had issued the INNs of 9,125 drugs in six languages (English, Latin, French, Russian, Spanish, Arabic, and Chinese). Since our targeted PI is written in French or English, we constructed a drug dictionary with INNs to use both languages for text mining. We further added Drugbank synonyms and brand names to our dictionary.

To identify characteristics of DDIs, we grouped drugs hierarchically into pharmacological and chemical subgroups and then according to the Anatomical Therapeutic Chemical classification system (ATC classification system), as shown in Figure 2. ATC codes, which the WHO develops and maintains, are used for the

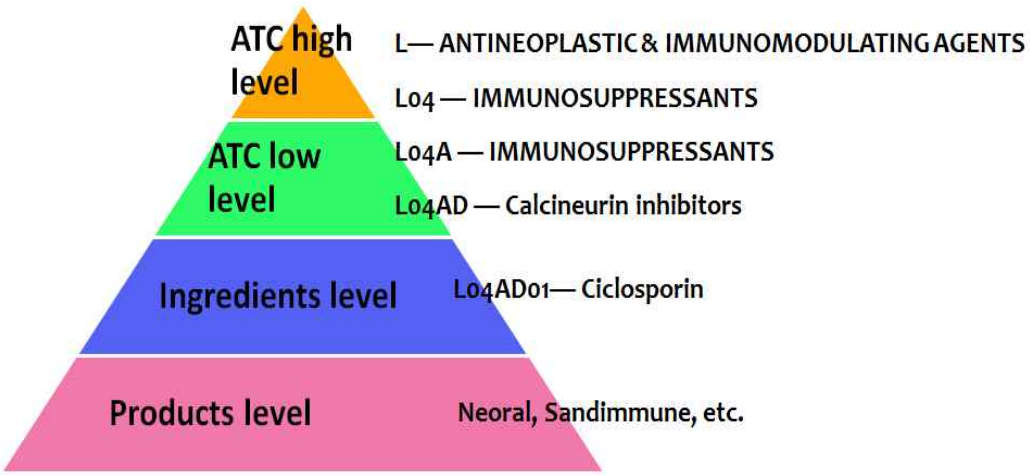

Figure 2 The drug hierarchy with ATC code

classification of drugs into five levels according to their therapeutic and chemical properties [15]. The levels represent the anatomical main group, the therapeutic main group, the therapeutic/pharmacological subgroup, the chemical/therapeutic/pharmacological subgroup, and the chemical substance. For example, fifth-level L04AD01 (ciclosporin) is further mapped into L (antineoplastic and immunomodulating agents) in the first level, L04 and L04A (immunosuppressants) in the second and third level, and L04AD (calcineurin inhibitors) in the fourth level.

In this study, PI is merged into the active ingredients and hierarchical groups with ATC codes to formulate characteristics of drug-drug interactions. In addition, as we used a fourth-level ATC code as the name of the drug class to extract implicit DDIs, we

collected two versions of ATC codes written in English and French. Especially the fourth level of ATC codes, which are ascribed therapeutic properties by using a word such as inhibitor or inducer, as in the cases of MAO inhibitors and HMG CoA reductase inhibitors, are used in the implicit extraction.

2.3 Drug-Drug interaction severity

Drug interactions are described mostly as drug names or drug class names in the three sections (Contraindications, Warnings and Precautions, and Drug interaction) of the PI. The Contraindications section is intended to describe known hazards and situations in which the risk from use clearly outweighs therapeutic benefit [16]; this should be referenced that medical experts should consult this section first when considering suitable drugs to administer. The Warnings and Precautions section includes a discrete set of adverse reactions for each prescription, and the Drug Interaction section is intended to identify potential drug interactions.

In this study, we targeted the three sections of the PI mentioned above for extracting DDIs and dividing them into two levels of DDI

severity. We defined DDIs extracted from the Contraindications section as adverse DDIs and those from the Warnings and Precautions and Drug Interaction sections as precautionary DDIs. We believe that the Contraindications section should refer to adverse DDIs because of its objective to highlight hazardous situations for medical experts. In many cases, the Drug Interaction section is included in a subsection of the Warnings and Precautions section of PI in the United States and Canada, but KEGG Drug classifies DDIs into two classes (CI and WP) according to their severity. Therefore, we considered that the CI of DDIs from KEGG Drug is mapped to our adverse DDIs, while WP from KEGG Drug is mapped to our Precautions DDI. The criteria used to classify DDI severity in the countries studied, excluding Japan, have not been decided based on our personal opinions but on the basis of PI sections authenticated by the national authorities.

2.4 Natural language extraction

To compare the similarity of the adverse DDI networks of the six countries, the PI of drugs with different trade names were merged to

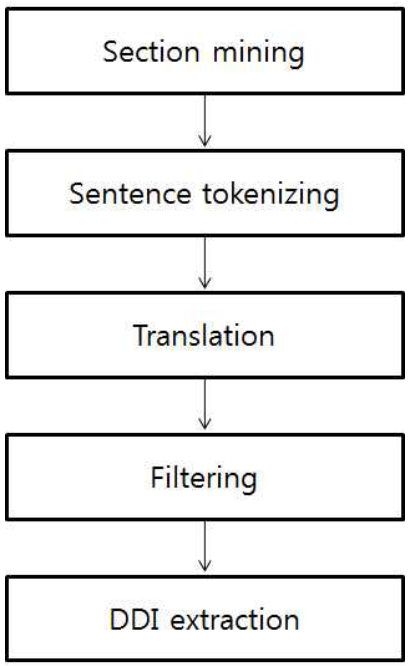

Figure 3 An overview of our approach in extracting drug interactions from prescribing information.

active ingredients and mapped to the drug dictionary. For this analysis, we targeted PI with a single active ingredient and then defined interactions between this active ingredient and other drug substances extracted from three sections of the PI. As the PI of France and Switzerland is written in French, we applied natural language processing (NLP) in two ways: first to the original French version and then to the translated English version.

Figure 3 shows the overall flow of natural language extraction. Three sections (Contraindications, Warnings and Precautions, and

Drug Interaction) that describe drug interactions were extracted from the documents. Analyzing the content of PDFs is challenging, as a PDF describes only the coordinates of pieces of text but not the structured form of a machine-readable method of text mining. After sentence tokenizing of the extracted sections, the information was translated into English using Google Translate API. Sentences that contain words and phrases such as “hypersensitive” and “allergic to” were interpreted as warnings related to a history of sensitive reactions to the similar components of the drug and thus were filtered from our target sentences for mining.

After filtering, we identified DDIs for explicit and implicit extraction from the text. Explicit extraction of a DDI described in a sentence as a drug name was achieved using the drug substance dictionary we constructed. Implicit extraction of a DDI referred to as a drug class name in a sentence was achieved by using the fourth level of ATC codes, and these extractions were converted to corresponding individual drug names (a fifth level of ATC codes). For example, selegiline should not be prescribed for patients who are being treated with antidepressant drugs, including MAO inhibitors, tricyclic antidepressants, and selective serotonin reuptake inhibitors

(e.g., citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, and sertraline). Citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, and sertraline are individual drug names that directly correspond to our drug dictionary, but the terms MAO inhibitors and selective serotonin reuptake inhibitors are drug class names that were converted to parglyline, alaproclate, zimelidine, and etoperidone.

2.5 iDrug

As shown in Figure 4, we constructed the relational iDrug database with DDIs extracted from the PI of the six countries by applying text-mining techniques. iDrug is composed of five tables, including a label table that stores the original text of the three sections of the PI from five countries, a drug table that contains meta information such as company and marketed status, and an ATC code table and interaction table storing the extracted DDIs from the six countries.

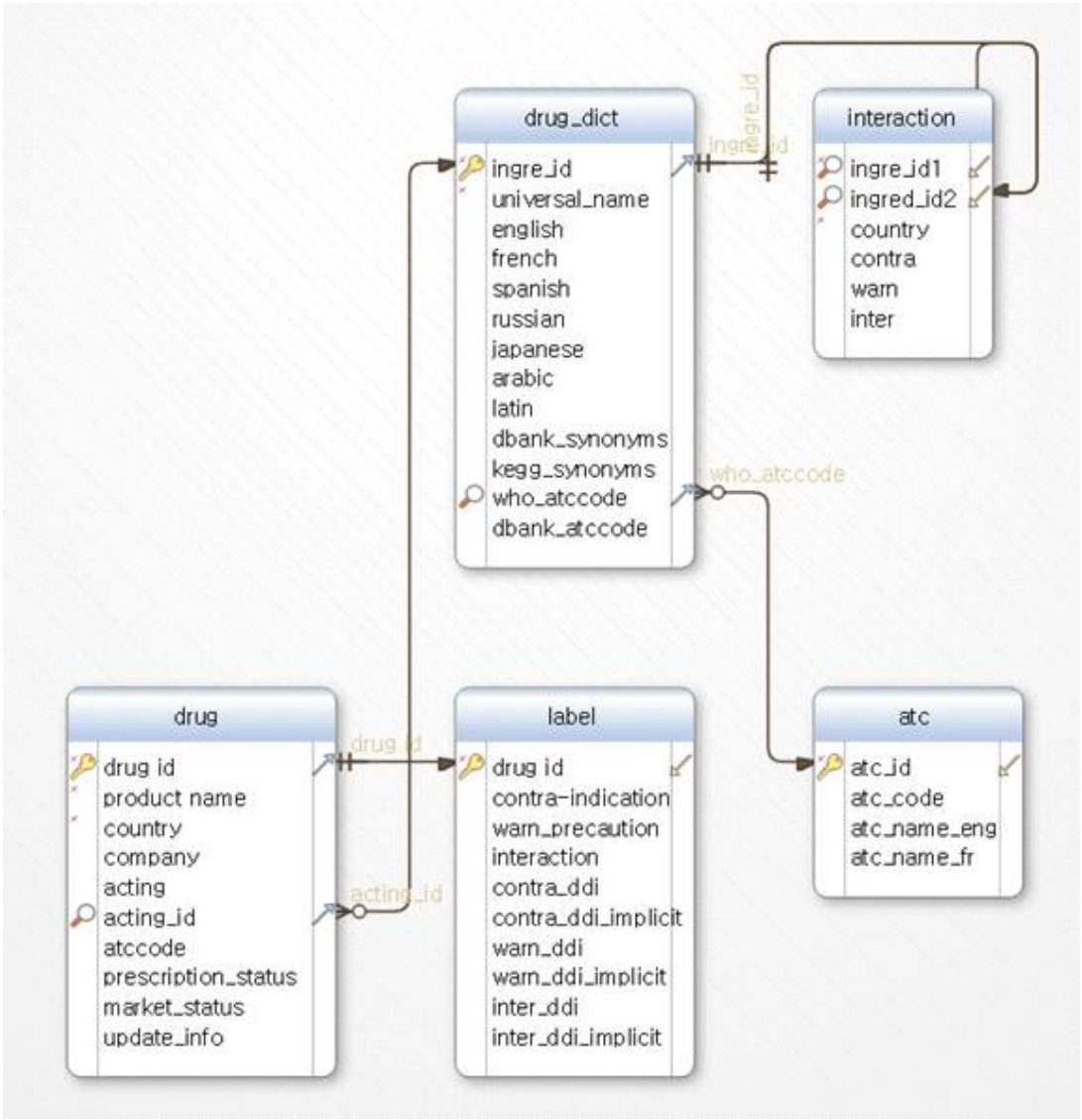

Figure 4 iDrug database schema

2.6 Analysis of International Adverse Drug-Drug Interaction Network

We created undirected adverse DDI networks with mining results from the Contraindications section of the PI from the six countries. In this network, node represents drug substance and edge represents interaction. To understand the inconsistency levels of the adverse DDIs of each country compared to the other countries, we created binary adjacency matrixes that interact with 1. Similarity between the adverse DDI networks of the six countries was calculated in two ways. First, the similarity score between two adverse DDI networks was defined as the binary Jaccard coefficient [17] as follows:

$$J(X,X') = \frac{|X \cap X'|}{|X \cup X'|} \quad (1)$$

where each nation binary vector X is represented by a profile $X=(x_{1 \text{ to } 1}, x_{1 \text{ to } 2}, x_{1 \text{ to } 3}, \cdots, x_{\text{last node to last node}})$ in which 1 or 0 based on whether each nation has defined interaction among all pairs of nodes. We computed all pairwise similarity of adverse DDIs between nations using the Jaccard coefficient. Second, the consistency level of adverse

DDIs between the six countries was calculated using the intraclass correlation coefficient [18]. In addition, to normalize the number of targeted PI units used for text mining, we selected 77 drugs sold by more than three companies. We analyzed the consistency level of adverse DDIs between five countries with a normalized number of targeted PI units, which is 231 per nation (77 drugs with 3 different brand names).

Ⅲ. Results and Discussion

Figure 5 shows the results of crawling PI from the representative drug compendia of five different countries. For analysis, we used only drugs with human prescriptions that were marketed recently. Among them, the percentage of PI with a single active ingredient is 68% for the United States, 90% for the United Kingdom, 91% for Canada, 83% for Switzerland, and 89% for France. This can be viewed as an indication that the results of the analysis in this study generally reflect characteristics of drug information authenticated in a nation. Through mapping PI with a single active ingredient into our drug dictionary, we finally analyzed targeted PI units for 15,241 from the United States, 4,071 from the United Kingdom, 6,594 from Canada, 5,028 from Switzerland, and 5,606 from France. In the case of the United States, the number of PI units used in DDI extraction was approximately three times higher than in other countries. Table 2 shows the interactions obtained from the targeted PI of the six countries. Adverse DDIs were extracted from the Contraindications section, and all DDIs were extracted from the Contraindications, Warnings and Precautions, and Drug Interaction sections.

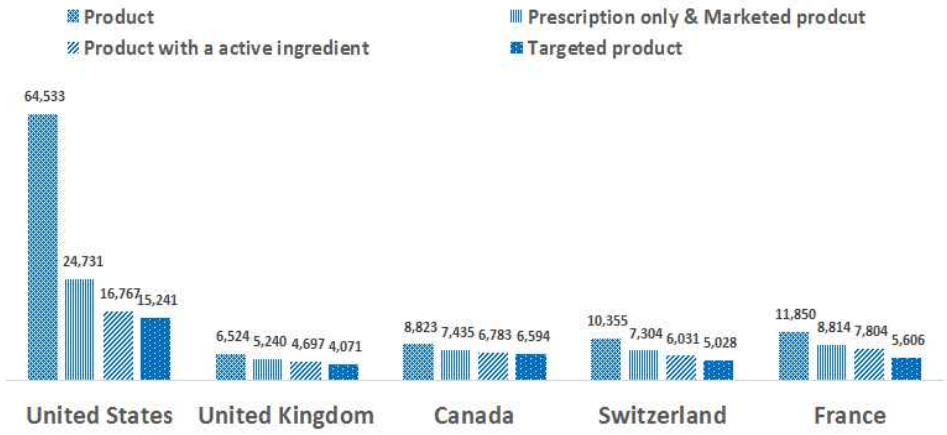

Figure 5 Compendia of PI from six countries for crawling and analysis

Table 2 Number of drug substance and drug interactions extracted from PI of six countries.

		United States	United Kingdom	Canada	Switzerland	France	Japan
Adverse DDIs	Drug	560	558	675	612	697	615
	Interaction	1,758	1,677	1,926	1,763	2,004	1,498
All DDIs	Drug	1,664	1,667	1,646	1,539	1,580	2,389
	Interaction	23,814	28,609	23,143	24,910	24,782	74,696

3.1 Evaluation

To assess the performance of iDrug, we defined Drugbank DDIs as the golden-standard (GS). We computed the overlap of the iDrug DDIs with the Drugbank GS, as depicted in Table 3. The overlap of explicit DDIs with Drugbank GS in all six countries was less than 50%. Particularly in the case of implicit DDIs, it showed a significantly low overlap with Drugbank GS. From this it can be inferred that Drugbank DDIs include only DDIs explicitly referred to as a drug names in the documents and exclude the association between the drugs and pharmacological subgroups referred to indirectly in the literature.

Furthermore, we found that it showed the lowest overlap of adverse DDIs between KEGG Drug and the Drugbank GS. This indicates that exiting drug databases provide different information related to DDIs. Therefore, the low overlap of iDrug and Drugbank GS can be explained with the constraints of the different drug testing environments and the absence of an integrated DDI database. In addition, the low overlap shows that Drugbank does not provide a large number of potential DDIs and highlights the importance of

formulating an integrated DDI severity database.

To further evaluate the performance of our approach in the extraction of DDIs, we randomly chose 100 PI units per nation and manually created a DDI golden-standard for each document. By comparing the adverse DDIs of iDrug and DDI GS manually created from the Contraindications sections of PI, F1 scores were computed, as shown in Table 4. For the implicit extraction, performance was measured based on the pharmacological subgroup names mentioned in the documents. We found the average F1 score of explicit DDIs to be 0.91, while the average F1 score of implicit DDIs is 0.47 due to 0.75 recall and 0.47 precision. We realized that most of the DDIs that remained undetected through text mining went undetected due to the limitations of our pharmacological subgroup dictionary. To compare adverse DDIs between the six countries using equal criteria, we used the fourth level of ATC codes as drug class names because they provide code names in both French and English. This approach shows a limitation in terms of the abundance of drug class names.

Table 3 Overlap of iDrug DDIs with Drugbank golden-standard (n=the number of interactions that match with Drugbank/ the number of extracted interactions)

		United States, % (n)	United Kingdom, % (n)	Canada, % (n)	Switzerland, % (n)	France, % (n)	Japan, % (n)
Adverse DDIs	Explicit DDIs	46.7 (315/751)	48.1 (390/811)	45.4 (373/821)	45.2 (353/781)	35.2 (379/1,077)	17.8 (266/1,498)
	Implicit DDIs	13.2 (288/2,179)	20.3 (202/997)	19.9 (242/1,216)	14.8 (159/1,076)	16.6 (177/1,064)	
DDIs		16.3 (3,874/23,814)	16.3 (4,664/28,609)	15.6 (3,612/24,143)	15.1 (3,768/24,910)	13.3 (3,307/24,782)	8.0 (6,006/74,696)

Table 4 Performance of the extracted interactions compared to manually created golden-standard

	United States		United Kingdom		Canada		Switzerland		France	
	Explicit DDI	Implicit DDI	Explicit DDI	Implicit DDI	Explicit DDI	Implicit DDI	Explicit DDI	Implicit DDI	Explicit DDI	Implicit DDI
Precision	0.94	0.54	0.93	0.51	0.94	0.53	0.93	0.41	0.92	0.38
Recall	0.92	0.85	0.92	0.76	0.86	0.69	0.88	0.80	0.92	0.75
F1 score	0.93	0.66	0.93	0.61	0.90	0.60	0.90	0.54	0.92	0.50

3.2 Characterization of International Adverse Drug-Drug Interactions

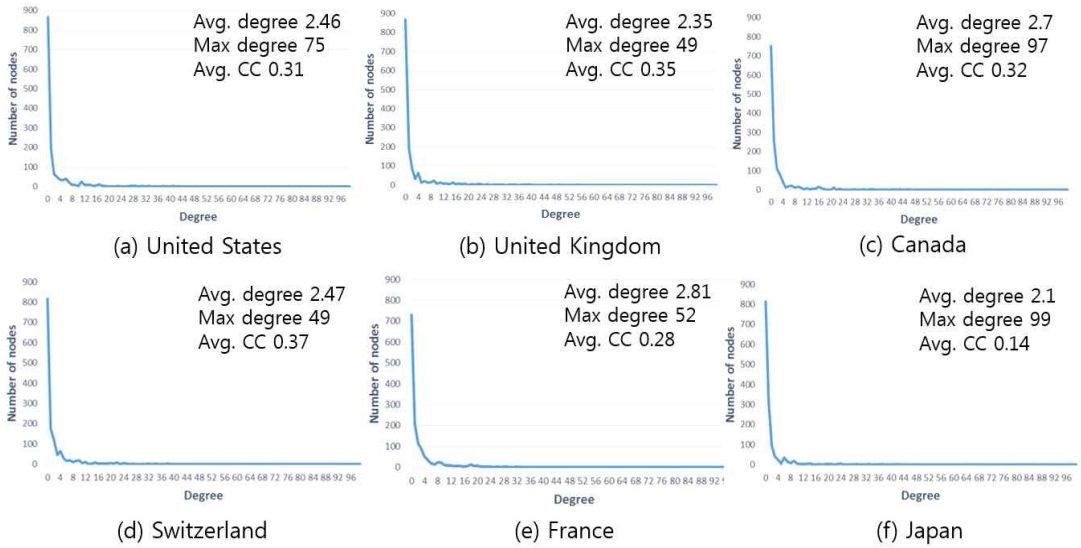

Figure 6 Degree distribution of international adverse DDIs

The degree distribution of the six countries followed power law distributions, as shown in Figure 6. These results suggest that a few drug nodes act as hubs with significantly large numbers of contraindications to other drugs and that most drug nodes have few contraindications. Isocarboxazid, an irreversible monoamine oxidase inhibitor for the treatment of depression, is contraindicated to 75 drug substances in the United States, and the anticoagulant acenocoumarol is contraindicated to 96 drug substances in Canada.

That the average clustering coefficient is 0.31 for the United States, 0.35 for the United Kingdom, 0.32 for Canada, 0.37 for

Switzerland, 0.24 for France, and 0.14 for Japan indicates that the adverse DDI networks of the six countries are connected sparsely. The adverse DDI networks for the six countries are clustered into the first level of ATC codes, as shown in Figure 7. Moreover, the number of nodes and edges used to create adverse DDI networks differ for each country. From the networks, we found that the majority of the countries have numerous super nodes in anti-infective drugs for systemic use (ATC code:J). Because J-class drugs such as HIV protease inhibitors and drugs for lepra/tuberculosis are associated with metabolic disease and development disease, we realized this class contains a high level of contraindication to avoid potential adverse drug reactions in people suffering from these types of diseases. Considering the top five highest degree drug nodes for each nation, anti-infective drugs for systemic use (ATC code:J) such as ritonavir, tipranavir, and voriconazole were selected.

3.3 Consistency of International Adverse Drug-Drug Interactions

To compare adverse DDI networks between the six countries at a glance, we created networks by fixing the position of nodes across

all networks with 1,365 union nodes, as shown in Figure 8. We found that the adverse DDIs of all six countries differ considerably. The results of pairwise Jaccard similarity coefficient are described in Figure 9. The majority of countries showed similarity coefficients below 0.6; in other words, the co-administration of two drugs is not contraindicated in the United States, but it is in Japan or Switzerland. These results highlight the importance of constructing an integrated DDI severity database using authoritative information such as officially approved PI. Although the number of targeted PI units in the United States is almost three times greater than those in the United Kingdom, this analysis gives great insight into aspects of comparing the characteristics of all drugs sold in a country.

The intraclass correlation coefficient (ICC) of 0.27 indicates a weak correlation between the adverse DDIs of the six countries [19]. Table 5 shows adverse DDIs listings in the six countries; only 1% of the adverse DDIs were listed in all six countries, and 3% were listed in five countries. After normalizing the number of targeted PI units per nation, the results of the consistency levels in the listing of adverse DDIs between the six countries were obtained (Figure 9). We found that similarity scores of the normalized number of drugs are

elevated above those of the entire targeted PI group but it still shows poor agreement [20].

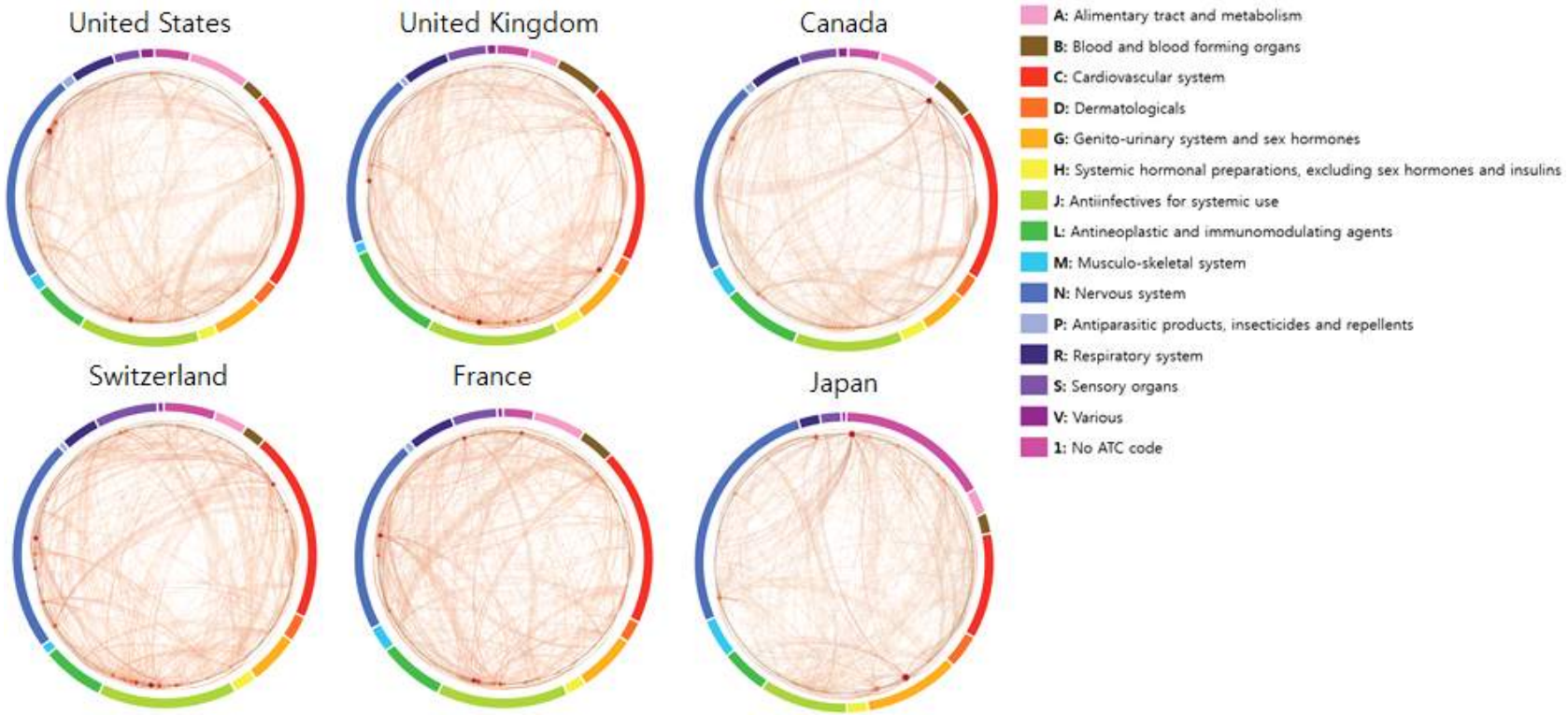

Figure 7 Adverse DDI networks (contraindications for coadministration) for six countries with first level of ATC code (Networks represent drugs and interactions extracted from each country's PI)

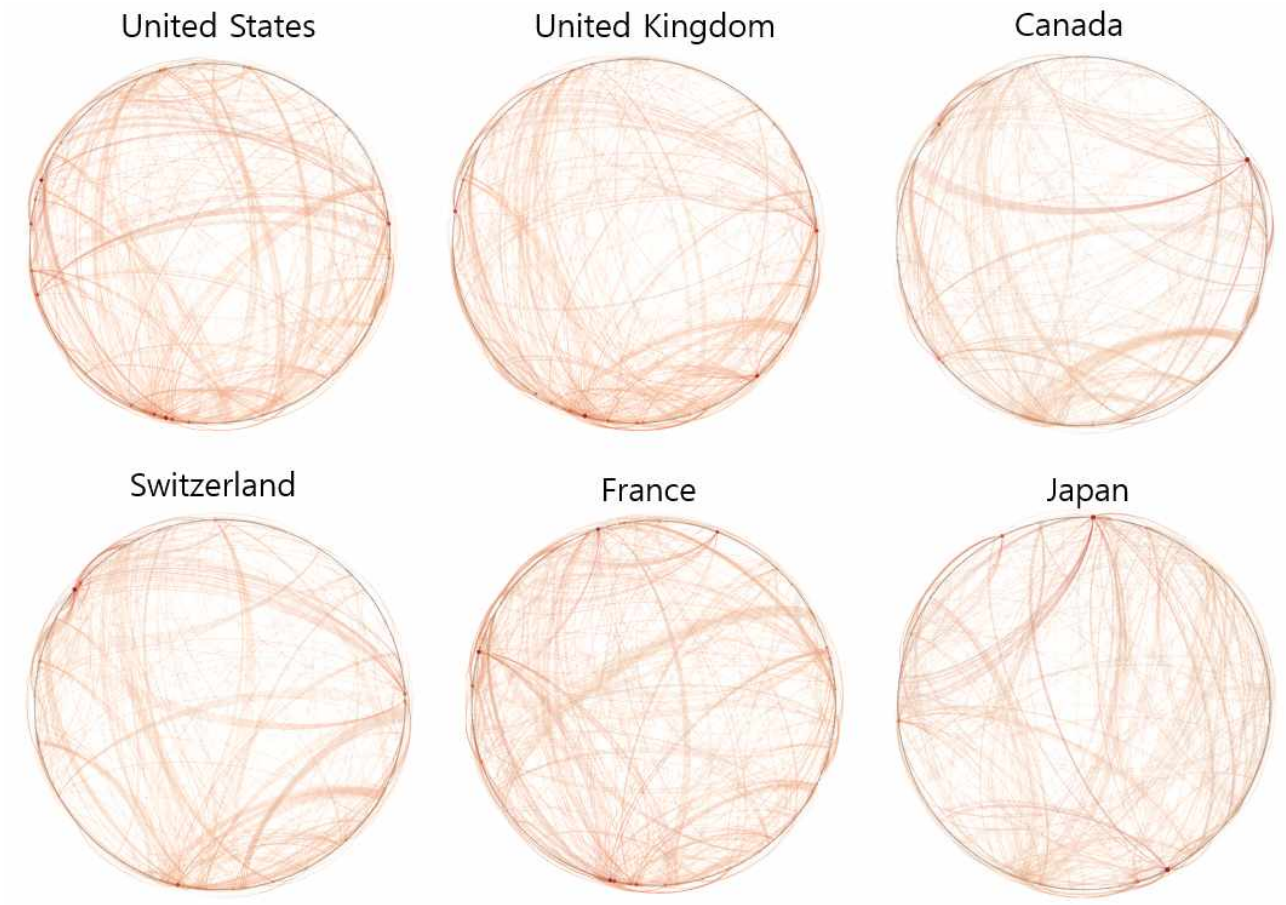

Figure 8 Adverse DDI networks (contraindications for coadministration) for six countries (Networks represent drugs that extracted from any six countries and interactions based on each country's PI)

Figure 9 Pairwise similarity of Adverse DDI network(Text mining driven adverse DDIs from the entire targeted PI group of each country(left) and PI with normalized 77 drug per each country(right).)

Table 5 Consistency between compendia of six countries in listing of adverse DDIs

	All six countries	Five countries	Four countries	Three countries	Two countries	One countries	Total
Number of adverse DDIs listed in different combination of countires	59	197	314	459	1,001	4,652	13,364

3.4 Community Detection of Drug-Drug Interactions

We clustered all DDIs extracted from the Contraindications, Warnings and Precautions, and Drug Interaction sections of the PI crawled from the United States, Dailymed, with a modularity. Figure 10 shows the results of community detection of DDIs in the United States with the first level of ATC codes. It created 12 communities with higher interlinks between nodes of the same community. We had expected one or two specific first level ATC codes would encompass a large portion of one community. In contrast, numerous first level ATC codes are mixed in one community. Nevertheless, when applying community detection with a modularity to the other five countries, we determined that from 11 to 13 common communities were detected from all five countries. These results indicate that DDI communities are clustered in the effects of significantly important biological pathways or main target proteins. For future work, by adding drug-target information to iDrug, we expect to discovery more biological characteristic of DDIs from the six countries.

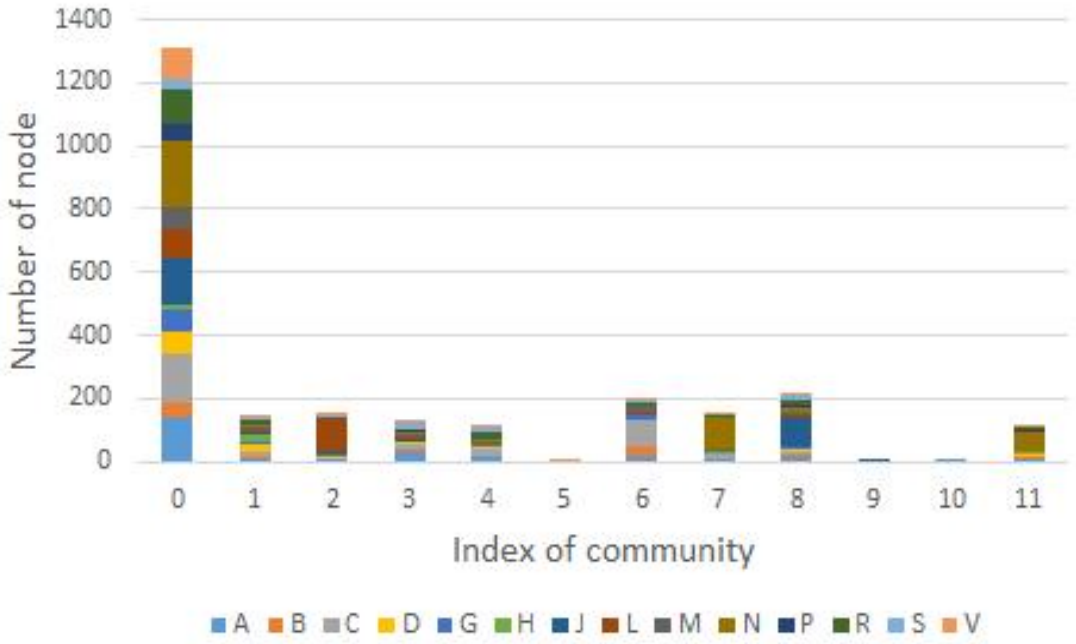

Figure 10 Community detection of DDIs from United States in iDrug

Ⅳ. CONCLUSION

In this study, we constructed an integrated DDI severity database by applying text mining techniques to the prescribing information of six countries (the United States, the United Kingdom, Canada, France, Switzerland, and Japan). Furthermore, we analyzed the characteristics of the DDI networks and consistency levels of the data from each country. We computed the pairwise differences of adverse DDIs between nations to understand the inconsistency levels of the DDIs of each country compared to the other countries. Our analysis shows that the similarity of adverse DDIs between countries is lower than 0.6. In other words, the co-administration of two drugs is not contraindicated in some countries, while in others it is. In addition, the intraclass correlation coefficient (ICC) of 0.27 indicates a weak correlation between the adverse DDIs of the six countries.

To the best of our knowledge, our work is the first attempt to automatically extract PI from the national regulatory authorities of the six countries by using text-mining techniques. Since our data were populated from PI approved by national regulatory authorities, we believe that our database can serve as a more authentic and reliable

guide for medical experts.

References

[1] Becker, Matthijs L., Kallewaard Marjon, Caspers Peter W.J., et al. "Hospitalisations and emergency department visits due to drug-drug interactions: a literature review." *Pharmacoepidemiology and Drug Safety* 16.6 (2007): 641.

[2] Van Der Sijs, Heleen, Vulto MSc, Arnold, Berg M.D., Marc. "Overriding of drug safety alerts in computerized physician order entry." *Journal of the American Medical Informatics Association* 13.2 (2006): 138-147.

[3] Cash, Jared J. "Alert fatigue." *American Journal of Health-System Pharmacy* 66.23 (2009).

[4] DRUG-REAX System [internet database]. Greenwood Village, CO: Thomson Reuters (Healthcare), 2009.

[5] Lexi-Comp OnlineTM, Lexi-Drugs Online™, Hudson, Ohio: Lexi-Comp, Inc. (2011).

[6] http://www.va.gov/

[7] Hansten, P.D. & Horn, J.R. Drug Interactions: Analysis and Management (Facts and Comparisons, St Louis, MO, 2007).

[8] Olvey, E. L., S. Clauschee, and D. C. Malone. "Comparison of critical drug - drug interaction listings: the Department of Veterans Affairs medical system and standard reference compendia." *Clinical Pharmacology & Therapeutics*87.1 (2010): 48-51.

[9] Wang, Lorraine M., Wong, Maple, Lightwood, James M., et al. "Black box warning contraindicated comedications: concordance among three major drug interaction screening programs." *Annals of Pharmacotherapy* 44.1 (2010): 28-34.

[10] Wishart, David S., Knox, Craig, Guo, An Chi, et al. "DrugBank: a comprehensive resource for in silico drug discovery and exploration." *Nucleic Acids Research* 34.suppl 1 (2006): D668-D672.

[11] Kanehisa, Minoru, Goto, Susumu, Sato, Yoko, et al. "KEGG for integration and interpretation of large-scale molecular data sets." *Nucleic Acids Research* (2011): gkr988.

[12] Takarabe, Masataka, Shigemizu, Daichi, Kotera, Masaaki, et al. "Characterization and classification of adverse drug interactions." *Genome Inform.* Vol. 22. 2010.

[13] Zineh, I., Gerhard, T., Aquilante C. L., et al. "Availability of pharmacogenomics-based prescribing information in drug package

inserts for currently approved drugs." *The Pharmacogenomics Journal* 4.6 (2004): 354-358.

[14] http://www.who.int/medicines/services/inn/en/

[15] http://www.whocc.no/atc_ddd_index/

[16] https://en.wikipedia.org/wiki/Package_insert

[17] Jaccard, Paul. "The distribution of the flora in the alpine zone." *New phytologist*11.2 (1912): 37-50.

[18] Koch, Gary G. "Intraclass correlation coefficient." *Encyclopedia of Statistical Sciences* (1982).

[19] Futrell, David. "When quality is a matter of taste, use reliability indexes."*Quality Progress* 28.5 (1995): 81-88.

[20] Nikolić, Božana S., and Maja S. Ilić. "Assessment of the consistency among three drug compendia in listing and ranking of drug-drug interactions." *Bosnian Journal of Basic Medical Sciences* 13.4 (2013): 253.

Abstract

iDrug: Text mining driven international adverse drug-drug interactions

Saejung Kim

Electrical and Computer Engineering

The Graduate School

Seoul National University

With the remarkable development of medical analysis techniques, many drug-drug interactions (DDI) have been identified. Especially, as drugs with adverse DDIs can bring severe side-effects when used together, this information should be given with priority to medical experts in an emergency situation such as in intensive care unit (ICU). However, drug compendia that provide drug-drug interaction severity often have inconsistencies of DDI severity among different compendia. Therefore, it is critical to identify the correct DDI severity and organize such information in a consistent and structured way.

We applied text mining techniques to the extraction of drug-drug interactions from prescribing information in six countries (the United States, the United Kingdom, Canada, France, Switzerland, and Japan), and constructed a DDI severity database that stores two levels of

DDI severity, contraindications and precautions, for the six countries. Based on the information we extracted, we analyzed the characteristics of the DDI network for each nation. We computed the pairwise differences of adverse DDIs between nations to understand the inconsistency levels of the DDIs of each country compared to the other countries. To the best of our knowledge, our work is the first attempt to automatically extract PI from the national regulatory authorities of the six countries by using text-mining techniques. Since our data were populated from PI approved by national regulatory authorities, we believe that our database can serve as a more authentic and reliable guide for medical experts.

keywords : Adverse drug-drug interaction, Text mining
***Student Number* : 2014-21735**