Arh. farm. 2022; 72: 689 – 700 *Original scientific paper/Originalni naučni rad*

Application of text-mining techniques for extraction and analysis of paracetamol and ibuprofen marketed products' qualitative composition

Jelena Djuriš*, Jovana Pilović, Marina Džunić, Sandra Cvijić, Svetlana Ibrić

University of Belgrade – Faculty of Pharmacy, Department of Pharmaceutical Technology and Cosmetology, Vojvode Stepe 450, 11221 Belgrade, Serbia

*Corresponding author: Jelena Djuriš, E-mail: jelena.djuris@pharmacy.bg.ac.rs

Abstract

Text mining (TM) applications in the field of biomedicine are gaining great interest. TM tools can facilitate formulation development by analyzing textual information from patent databases, scientific articles, summary of products characteristics, etc. The aim of this study was to utilize TM tools to perform qualitative analysis of paracetamol (PAR) and ibuprofen (IBU) formulations, in terms of identifying and evaluating the presence of excipients specific to the active pharmaceutical ingredient (API) and/or dosage form. A total of 152 products were analyzed. Web-scraping was used to retrieve the data, and Python-based open-source software Orange 3.31.1 was used for TM and statistical analysis (ANOVA) of the obtained results. The majority of marketed products for both APIs were tablets. The predominant excipients in all tablet formulations were povidone, starch, microcrystalline cellulose and hypromellose. Povidone, stearic acid, potassium sorbate, maize starch and pregelatinized starch occurred more frequently in PAR tablets. On the other hand, titanium dioxide, lactose, shellac, sucrose and ammonium hydroxide were specific to IBU tablets. PAR oral suspensions more frequently contained dispersible cellulose; liquid sorbitol; methyl and propyl parahydroxybenzoate, glycerol and acesulfame potassium. Specific excipients in other PAR dosage forms, such as effervescent tablets, hard capsules, oral powders, solutions and suspensions, as well as IBU gels and soft capsules, were also evaluated.

Key words: text mining, dosage forms, qualitative analysis, excipients, paracetamol, ibuprofen

https://doi.org/10.5937/arhfarm72-40397

Introduction

Text mining (TM) is a process of analyzing unstructured textual data to discover patterns and gain new knowledge and insights related to the studied topic. It requires the combined application of natural language processing (NLP), data science, statistics and/or machine learning. One of the areas with great potential for TM techniques as research tools is biomedicine. TM is used to retrieve chemical drug names, structures, associated properties, disease names, gene and protein names, clinical textual data, and other relevant information for drug discovery (1), drug repurposing (2-3) and many other applications (4-6).

TM is also extensively applied in the development of new products for various applications. Choi and Hong (7) have proposed TM of patent databases as one of the efficient methods to search for tailored-based sophisticated technologies such as nanomedicines. Patel et al. (8) have utilized TM to explore the patent database on probiotic formulations. The food industry has recognized the great potential of TM by utilizing it to evaluate dietary patterns through an analysis of web searches, surveys, social media information etc. (9). TM of scientific articles allows understanding of the state-of-the-art data on a specific topic, which was demonstrated by the study of microalgal paste and powder as food and feed (10). TM also enables mining of consumer opinions on specific products, highly facilitating the development of new products (9). Specific machine learning algorithms can use databases of culinary recipes and create a recommendation system for giving suggestions tailored to consumer requirements (11). Such an approach was also used to mine the data on cosmetic products formulations (12). Authors (12) have used miscellaneous sources to retrieve the formulation data, including scientific articles, handbooks, and websites. Their recommendation system aids in selection of formulation ingredients, including quantitative composition, and also provides information on the processing steps required to produce the cosmetic product. Kawi et al. (13) have proposed an interesting framework for recommendation of personalized pharmaceutical formulations, based on text retrieval and processing.

It is evident that TM of formulation compositions can be a valuable tool for the development of new pharmaceutical products. Rincón-López et al. (14) have explored patent databases using TM tools in order to review the technological evolution of cyclodextrins as excipients in pharmaceutical products. The same authors (15) have further exploited the network science to examine the patenting trends and roles of cyclodextrins in specific dosage forms and formulation types, such as tablets, powders, suspensions, emulsions, etc. An insightful study on TM application for analysis of drug-formulation based recalls in the USA was also recently published (16).

The aim of this study was to test whether the TM approach can differentiate between the dosage forms and formulations of two active pharmaceutical ingredients (APIs) that are similar in terms of pharmacotherapy, yet differ in physico-chemical and pharmaceutical-technical properties. The selected APIs were paracetamol (PAR) and ibuprofen (IBU). Both are used predominantly as analgesics in a variety of dosage forms.

The analysis of the global market share of oral over-the-counter (OTC) analgesics reveals that paracetamol is the predominant API, with almost one third of the market share (27.4% in 2019 and estimated 27.1% in 2027) (17). Nonsteroidal anti-inflammatory drugs (NSAID) are the major group of analgesics on the market (36.2% in 2019 and estimated 35.7% in 2022), followed by salicylates (~19%) and others (~10%) (17). Ibuprofen is one of the key NSAIDs, in terms of the presence of its products on the market (18).

Experimental part

Methods

Marketed PAR and IBU products were searched on *Electronic Medicines Compendium* (EMC) website in 2021 (19). This website enables a simple and reliable search of marketed pharmaceuticals, approved by the UK MHRA (*Medicines and Healthcare Products Regulatory Agency*) and, historically, by the EMA (*European Medicines Agency*). From the obtained results, Summaries of Products Characteristics (SmPCs) for 152 products were further analyzed, to reflect the differences in dosage forms and formulations. A majority of the analyzed formulations were PAR products (111), whereas additional 41 IBU products were used to evaluate the potential influence of the active pharmaceutical ingredient on the formulation composition. The ratio between the analyzed PAR and IBU products reflects their relative presence on the market.

In order to collect the list of excipients and dosage form for each product, a freely available web-scraper *ParseHub* (20) was used. Once the desired information from a website is introduced to the web-scraper in the form of a project, the results are obtained as a CSV database. Python-based open-source software Orange 3.31.1 was used for the text mining and statistical analysis (ANOVA) of the obtained results. Web-scraped CSV database of the selected formulations was used as the text corpus for the TM procedure. The following text transformations were used prior to generating word clouds: lowercase transformation, specific Regexp tokenization and filtering, in order to recognize both one and two (or multiple) word excipients that were comma separated in the excipient list. Once the text was preprocessed, word counts and weights were generated from tokens and further evaluated.

Some of the figures were prepared in the Jupyter Notebook environment using Python 3.8.1.

Results and discussion

Solid dosage forms

Figure 1 represents the distribution of the dosage forms for the studied products (PAR and IBU). The majority of marketed products for both APIs are tablets (including both coated and uncoated ones). In the case of PAR, the most frequent dosage form were hard capsules, followed by oral powders (this category includes powders for oral solution

as well) and effervescent tablets. PAR products are predominantly solid dosage forms. IBU products, on the other hand, appear in liquid and semi-solid forms as well.

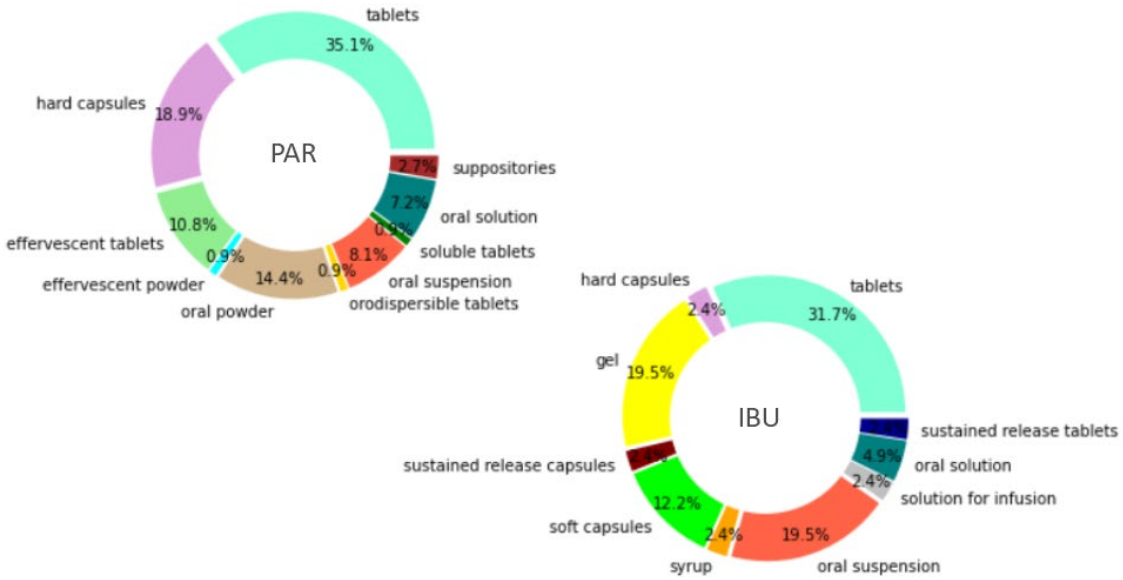

Figure 1. Distribution of the dosage forms for the representative samples of paracetamol (PAR) and ibuprofen (IBU) marketed products

Slika 1. Raspodela različitih farmaceutskih oblika lekova u reprezentativnom uzorku registrovanih preparata paracetamola (PAR) i ibuprofena (IBU)

Tablet formulations were first analyzed as the most frequent dosage form in the sample. Figure 2 represents the word cloud of the tablet formulations, where the size of each word in the cloud is proportional to its presence in the analyzed text, in this instance in the list of excipients. The main excipients in all tablet formulations were the typical excipients for this solid oral dosage form: povidone, starch (predominantly maize starch, including pregelatinized maize starch) microcrystalline cellulose and hypromellose. The role(s) of the selected excipients in solid pharmaceutical dosage forms, as well as the recommended concentrations for each role, are represented in Table I. Student's t-test was used to compare tablet formulations of PAR and IBU. It has been determined that there are some statistically significant differences in terms of the presence or absence of specific excipients in tablets. More precisely, the following excipients have occurred more frequently in PAR tablets compared to IBU tablets: povidone ($p=0.001$), stearic acid ($p=0.001$), potassium sorbate ($p=0.001$), maize starch ($p=0.013$) and pregelatinized starch ($p=0.036$). Likewise, the following excipients have more frequently occurred in IBU tablets compared to PAR tablets: titanium dioxide ($p=0.001$), lactose ($p=0.022$), shellac ($p=0.027$), sucrose ($p=0.033$) and ammonium hydroxide ($p=0.033$).

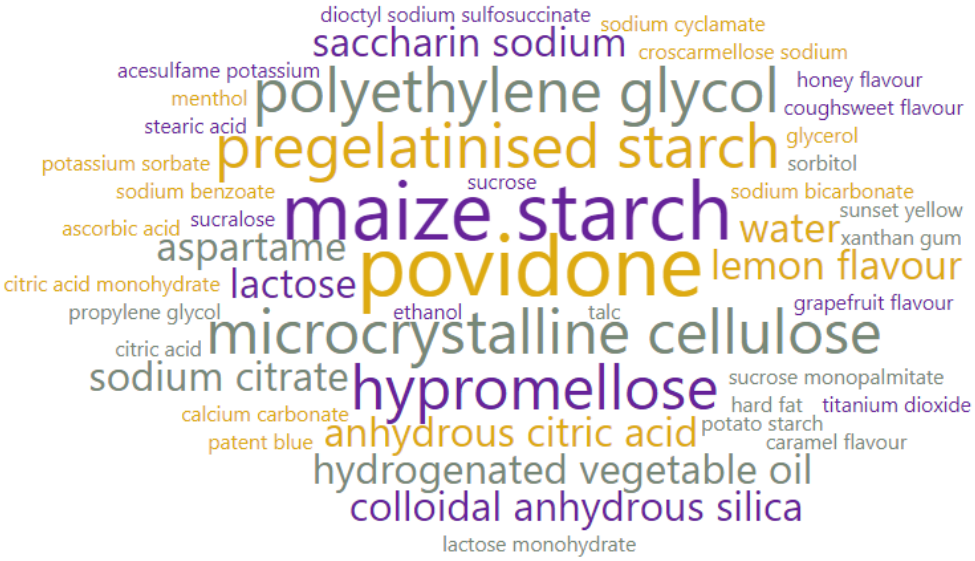

Figure 2. **Word cloud for the composition of tablet formulations (word size is proportional to its presence in the list of excipients)**

Slika 2. **Oblak reči za sastave formulacija tableta (veličina reči je srazmerna njenoj učestalosti u listi ekscipijenasa)**

Table I The role(s) of selected excipients in solid pharmaceutical dosage forms (25)

Tabela I Uloge odabranih ekscipijenasa u čvrstim farmaceutskim oblicima (25)

Excipient	Role	Concentration (%)
Povidone	Drug carrier Suspending/dispersing agent Tablet binder, diluent or coating agent	10 – 25 up to 5 0.5 – 5
Microcrystalline cellulose	Tablet diluent/binder Capsule diluent/binder Tablet disintegrant Antiadherent Adsorbent	20 – 90 20 – 90 5 – 15 5 – 20 20 – 90
(Maize) starch	Tablet/capsule diluent Tablet binder (starch paste) Tablet disintegrant	20 – 50 5 – 25 3 – 15
Pregelatinized (maize) starch	Capsule diluent Tablet binder Tablet disintegrant	5 – 75 5 – 20 5 – 10
Hypromellose	Tablet binder Coating agent Matrix forming material Thickening agent	2 – 5 2 – 20 10 – 80 0.45 – 1

The differences between the excipients used in formulations can be attributed to differences in physico-chemical properties of PAR and IBU, including mechanical properties such as compressibility and compactibility. PAR is well known for its poor flow and compaction behavior (21), with PAR tablet formulations often requiring wet or dry granulation prior to compaction. Although preservatives, such as potassium sorbate, are not often used in solid dosage forms, it is reported that such excipients are used to preserve the microbial quality of hygroscopic products and/or if wet granulation is used for processing with slurries or pastes (e.g. starch paste) that can be contaminated if stored prior to utilization in production (22). Several studies have indicated incompatibilities between IBU and boundary lubricants, in terms of identified changes in crystallinity and formation of eutectic mixtures of IBU and excipients such as stearic acid and magnesium stearate (23-24). Traditionally, IBU tablets are sugar (sucrose) coated in order to mask their unpleasant taste, as well as due to the acidity of IBU. Water protective coatings based on excipients such as shellac are often applied from non-aqueous solvents prior to application of sugar coats. Pharmaceutical glaze is a specially denatured alcoholic shellac solution that may contain titanium dioxide as an opacifying agent (25). Sugar coating is more frequently used for IBU compared to PAR tablets. Moreover, due to poor IBU solubility, it may be postulated that ammonium hydroxide was present in some IBU tablet formulations to increase drug solubility by interacting with polyethylene glycols that are partially ionized with hydroxide species, as suggested by Lodha et al. (26).

The differences in formulations could also be attributed to incompatibilities between the drug and specific excipients. Mazurek-Wadołkowska et al. (27) have utilized differential scanning calorimetry to investigate the potential for solid state interactions between common tablet excipients (povidone, crospovidone, microcrystalline cellulose, pregelatinized starch and magnesium stearate) and PAR. All the studied excipients, except crospovidone (Kollidon CL), have demonstrated compatibility with PAR.

In the analyzed sample, PAR effervescent tablets, as a specific dosage form, were also present. Povidone was used as binding agent in all PAR effervescent tablets formulations, as well as the combination of anhydrous citric acid as an acidic component and sodium hydrogen carbonate (sodium bicarbonate) and anhydrous sodium carbonate as a basic component of effervescent agents. It was reported that sodium bicarbonate increased PAR absorption (25). Another excipient which was present in almost all formulations is saccharin sodium, an intense sweetening agent (25). Half of the analyzed PAR effervescent tablets also contained aspartame, another intense sweetening agent (25). The other half of the analyzed PAR effervescent tablets contained sorbitol or sucralose and sucrose monopalmitate. Simethicone or dimethicone were used as antifoaming agents in 80% of the analyzed formulations. Polysorbate 80 or sodium lauryl sulfate as surface active agents (also acting as soluble lubricants in effervescent tablets formulations) were used in combination with simethicone or dimethicone, respectively. A few formulations contained polyethylene glycol as a lubricant, and only one formulation included a preservative (sodium benzoate).

Once the tablet formulations were analyzed, the next step was the analysis of other solid dosage forms. As represented in the first section, PAR formulations are predominantly solid dosage forms. Formulations of hard capsules and oral powders (including powders for oral solution) were analyzed and represented in Figure 3. PAR hard capsules formulations contain typical excipients, including gelatin for the capsule shell formation, titanium dioxide as a white pigment and opacifier, and magnesium stearate, primarily used as a lubricant in capsule and tablet manufacturing (25). Titanium dioxide has been recognized as a potentially genotoxic excipient and great efforts are being made to find a suitable alternative regarding tablet and/or capsule opacifiers (28). In addition, sodium lauryl sulphate, used as solubilizing agent and lubricant in tablets and capsules formulations (29), was identified in many formulations. Maize starch, used as diluent, and croscarmellose sodium and sodium starch glycolate, used as (super)disintegrants, were also frequent. Formulations of oral powders, on the other hand, are tailored to mask the unpleasant PAR taste. Oral PAR powders in the analyzed sample typically contain sucrose, sodium citrate, citric acid and lemon flavour. Some of the PAR oral powders had ascorbic acid as an additional API.

Figure 3. Word clouds for the composition of PAR hard capsules (left) and PAR oral powders (right)

Slika 3. Oblak reči za formulacije tvrdih kapsula PAR (levo) i oralnih praškova PAR (desno)

Liquid dosage forms

Peroral suspensions of PAR and IBU, as well as peroral solutions of PAR, were taken into consideration in terms of comparing their composition. Figure 4 represents the word clouds of these products' list of excipients. It is obvious from the word clouds that compositions of these formulations are different. If PAR solutions and suspensions are compared, the following excipients occur more frequently in PAR solutions compared to PAR suspensions: propylene glycol, sodium citrate, ethanol ($p=0.001$), levomenthol ($p=0.008$), sodium cyclamate ($p=0.025$) and sorbitol ($p=0.025$). Furthermore, the following excipients occurred more frequently in PAR suspensions compared to PAR solutions: dispersible cellulose, sorbitol liquid (70% non-crystallizing) ($p=0.002$),

polysorbate 80 (p=0.003), strawberry flavor (p=0.003), methyl parahydroxybenzoate (p=0.015) and glycerol (p=0.018). Propylene glycol is a viscous liquid, miscible with water and ethanol, frequently used as co-solvent in solutions for peroral administration. Sodium citrate is most often used with citric acid as a buffering agent to adjust the pH of solutions, or as a sequestering agent (25). Levomenthol is used as a flavoring agent and/or odor enhancer. Sodium cyclamate is used as an artificial sweetening agent.

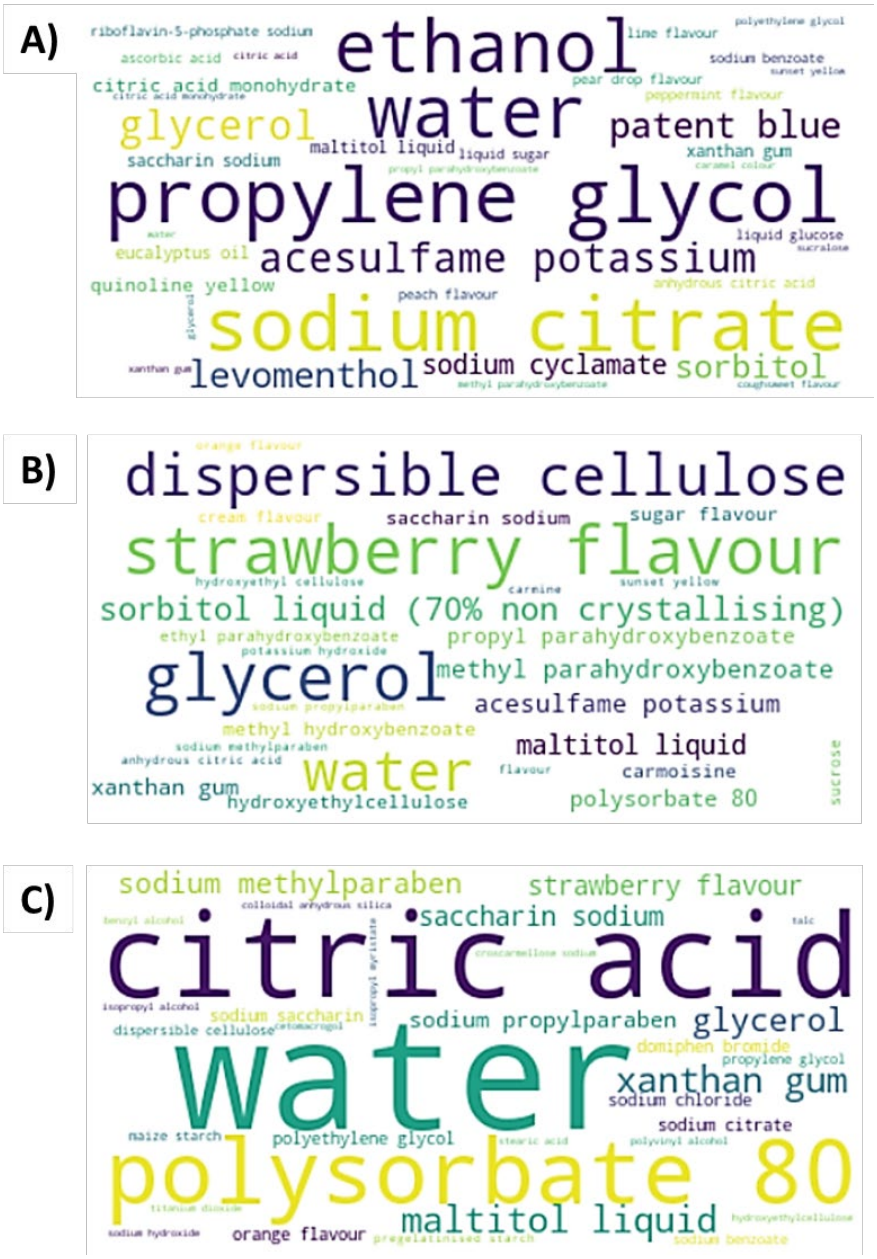

Figure 4. **Word clouds for the composition of A) PAR oral solutions, B) PAR oral suspensions and C) IBU oral suspensions**

Slika 4. **Oblaci reči za formulacije A) peroralnih rastvora PAR, B) peroralnih suspenzija PAR i C) peroralnih suspenzija IBU**

In order to evaluate the influence of the API on the dosage form composition, PAR and IBU oral suspensions' formulations were compared. The following excipients occurred more frequently in PAR suspensions compared to IBU suspensions: dispersible cellulose (p=0.001), sorbitol liquid (70% non-crystallizing) (p=0.002), methyl parahydroxybenzoate (p=0.003), propyl parahydroxybenzoate (p=0.028), glycerol (p=0.032) and acesulfame potassium (p=0.017). Dispersible cellulose is a mixture of microcrystalline cellulose and carboxymethylcellulose sodium that can be dispersed in water to form a thixotropic gel suitable to be used as suspending vehicle (25).

The only excipient that occurred more frequently in IBU suspensions compared to PAR suspensions is citric acid (p=0.032).

Semi-solid and other dosage forms

IBU was also formulated in semi-solid dosage forms, such as gels for topical administration. Two types of gel formulations were identified: 1) hydroxyethylcellulose gels with isopropyl alcohol used as a solvent for IBU, sodium hydroxide for pH adjustment and benzyl alcohol as preservative; 2) carbomer gels with propylene glycol used as a solvent for IBU, diethylamine for pH adjustment and isopropyl myristate as an IBU permeation enhancer (30). Gels and oral suspensions were equally distributed in the analyzed sample, followed by soft capsules. IBU soft capsules, including one chewable soft capsule, are relatively complex formulations. In addition to gelatin, IBU soft capsules contain sorbitol, polyethylene glycol 400, various lipid excipients (triglycerides, hard fat, etc.), glycerol, lecithin and other excipients, such as polysorbate 80 as a surfactant, or isopropyl alcohol as a co-solvent.

Conclusion

The study presented an investigation of PAR and IBU formulations, in terms of analyzing and comparing the qualitative composition of their different dosage forms. PAR and IBU are two representative NSAID medicines and, by utilizing text mining tools, it was possible to identify similarities and differences between the formulations. This study has confirmed the great potential of TM tools in the evaluation of pharmaceutical products' qualitative composition.

Acknowledgment

This study has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, contract number 451-03-68/2022-14/200161.

References

1. Zheng S, Dharssi S, Wu M, Li J, Lu Z. Text mining for drug discovery. In: Larson R, Oprea T, editors. Bioinformatics and Drug Discovery. Methods in Molecular Biology, vol. 1939. New York: Humana Press; 2019; p. 231-52.
2. Tari LB, Patel JH. Systematic drug repurposing through text mining. In: Kumar V, Tipney H, editors. Biomedical Literature Mining. Methods in Molecular Biology, vol. 1159. New York: Humana Press; 2014; p. 253-67.
3. Baker NC, Ekins S, Williams AJ, Tropsha A. A bibliometric review of drug repurposing. Drug Discov Today. 2018;23(3):661-72.
4. Conceição SI, Couto FM. Text Mining for Building Biomedical Networks Using Cancer as a Case Study. Biomolecules. 2021;11(10):1430.
5. Wang Y, Zhu S, Liu C, Deng H, Zhang Z. Text mining and hub gene network analysis of endometriosis. BioMed Res Int. 2021;2021:5517145.
6. Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J. 2020;18:1414-28.
7. Choi Y, Hong S. Qualitative and quantitative analysis of patent data in nanomedicine for bridging the gap between research activities and practical applications. World Patent Information. 2020;60:101943.
8. Patel S, Patel MS, Patel AD, Shah UH, Patel MM, Solanki ND, Patel MJ. Probiotic Formulations: A Patent Landscaping Using the Text Mining Approach. Curr Microbiol. 2022;79(5):1-15.
9. Tao D, Yang P, Feng H. Utilization of text mining as a big data analysis tool for food science and nutrition. Compr Rev Food Sci F. 2020;19(2):875-94.
10. Raja R, Coelho A, Hemaiswarya S, Kumar P, Carvalho IS, Alagarsamy A. Applications of microalgal paste and powder as food and feed: An update using text mining tool. Beni-Suef Univ J Basic App Sci. 2018;7(4):740-7.
11. Chai KH, Hang CC, Xie M, editors. A text mining-based recommendation system for customer decision making in online product customization. Proceedings of an IEEE International conference on management of innovation and technology; 2006 June 21-23; Singapore. New Jersey: IEEE, Piscataway; 2006. 473 p.
12. Mining heterogeneous data for formulation design. Proceedings of an IEEE International Conference on Data Mining Workshops; 2020 Nov 17-20; virtual conference. IEEE; 2020. 589 p.
13. Medical Formulation Recognition (MFR) using Deep Feature Learning and One Class SVM. Proceedings of an IEEE 2020 International Joint Conference on Neural Networks; 2020 July 19-24; Glasgow, United Kingdom. IEEE; 2020. 1 p.
14. Rincón-López J, Almanza-Arjona YC, Riascos AP, Rojas-Aguirre Y. Technological evolution of cyclodextrins in the pharmaceutical field. J Drug Deliv Sci Technol. 2021;61:102156.
15. Rincón-López J, Almanza-Arjona YC, Riascos AP, Rojas-Aguirre Y. When cyclodextrins met data science: Unveiling their pharmaceutical applications through network science and text-mining. Pharmaceutics. 2021;13(8):1297.
16. Yang M, Byrn SR, Clase KL. An analytic investigation of the drug formulation-based recalls in the USA: see more beyond the literal. AAPS PharmSciTech. 2020;21(5):1-10.

17. Oral OTC Analgesics Market Share Trends [Internet]. [cited 2022 Aug 20]. Available from: https://dataintelo.com/report/oral-otc-analgesics-market/.
18. Non-steroidal Anti-Inflammatory Drugs (NSAIDs) Market Research Report 2021-2028 [Internet]. [cited 2022 Aug 20]. Available from: https://dataintelo.com/report/non-steroidal-anti-inflammatory-drugs-nsaids-market/.
19. Electronic Medicines Compendium (EMC) [Internet]. [cited 2022 Jul 10]. Available from: https://www.medicines.org.uk/emc/.
20. ParseHub - web scraping software [Internet]. [cited 2022 Jul 10]. Available from: https://www.parsehub.com/.
21. Šimek M, Grünwaldová V, Kratochvíl B. Comparison of compression and material properties of differently shaped and sized paracetamols. KONA Powder Part J. 2017;34:197-206.
22. Niazi SK. Handbook of Pharmaceutical Manufacturing Formulations: Volume One, Compressed Solid Products. Boca Raton: CRC Press, Taylor & Francis Group; 2019.
23. Stojanovska Pecova M, Geskovski N, Petrushevski G, Chachorovska M, Krsteska L, Ugarkovic S, Makreski P. Solid-state interaction of ibuprofen with magnesium stearate and product characterization thereof. Drug Dev Ind Pharm. 2020;46(8):1308-17.
24. Chaiya P, Chinpaisal C, Limmatvapirat S, Phaechamud T. Alteration of crystallinity and thermal properties from incompatibility between ibuprofen and boundary lubricants. Mater Today-Proc. 2021;47:3500-8.
25. Rowe RC, Sheskey P, Quinn M. Handbook of Pharmaceutical Excipients. 6th edition. London: Pharmaceutical Press; 2009.
26. Lodha A, Patel A, Chaudhuri J, Jadia P, Joshi T, Dalal J. Formulation and evaluation of transparent ibuprofen soft gelatin capsule. J Pharm Bioall Sci. 2012;4(1):S95.
27. Mazurek-Wadołkowska E, Winnicka K, Czajkowska-Kośnik A, Czyzewska U, Miltyk W. Application of differential scanning calorimetry in evaluation of solid state interactions in tablets containing acetaminophen. Acta Pol Pharm. 2013;70(5):787-93.
28. Blundell R, Butterworth P, Charlier A, Daurio D, Degenhardt M, Harris D, et al. The role of titanium dioxide (E171) and the requirements for replacement materials in oral solid dosage forms: An IQ Consortium Working Group Review. J Pharm Sci. 2022. doi: 10.1016/j.xphs.2022.08.011.
29. Sodium laurilsulfate used as an excipient [Internet]. European Medicines Agency report [cited 2022 Aug 21]. Available from: https://www.ema.europa.eu/en/documents/report/sodium-laurilsulfate-used-excipient-report-published-support-questions-answers-sodium-laurilsulfate_en.pdf.
30. Lakshmi PK, Kumar MK, Sridharan A, Bhaskaran S. Formulation and evaluation of ibuprofen topical gel: a novel approach for penetration enhancement. Int J Appl Pharm. 2011;3(3):25-30.

Primena tehnika za sistematizovanu obradu tekstualnih informacija u cilju analize kvalitativnog sastava registrovanih preparata paracetamola i ibuprofena

Jelena Đuriš*, Jovana Pilović, Marina Džunić, Sandra Cvijić, Svetlana Ibrić

Univerzitet u Beogradu – Farmaceutski fakultet, Katedra za farmaceutsku tehnologiju i kozmetologiju, Vojvode Stepe 450, 11221 Beograd, Srbija

*Autor za korespondenciju: Jelena Đuriš, E-mail: jelena.djuris@pharmacy.bg.ac.rs

Kratak sadržaj

Primena *text mining* (TM) alata u oblasti biomedicine postaje sve značajnija. TM alati mogu da olakšaju razvoj formulacija, tako što omogućavaju analizu tekstualnih informacija iz patentnih baza, naučnih članaka, sažetaka karakteristika lekova, itd. Cilj ovog rada bila je primena TM alata za kvalitativnu analizu formulacija paracetamola (PAR) i ibuprofena (IBU), u smislu identifikacije i procene prisustva ekscipijenasa koji su karakteristični za lekovitu supstancu i/ili farmaceutski oblik. Ukupno je analiziran sastav 152 preparata. Web-scraping je primenjen za prikupljanje podataka, a Orange 3.31.1, softver otvorenog koda zasnovan na programskom jeziku Python, primenjen je za TM i statističku analizu (ANOVA) dobijenih rezultata. Većina analiziranih formulacija za obe lekovite supstance bile su tablete, a najzastupljeniji ekscipijensi u njima su bili povidon, skrob, mikrokristalna celuloza i hipromeloza. Povidon, stearinska kiselina, kalijum sorbat, kukuruzni skrob i pregelirani skrob se češće pronalaze u formulacijama PAR tableta. Titanijum-dioksid, laktoza, šelak, saharoza i amonijum hidroksid su specifični za IBU tablete. PAR peroralne suspenzije su češće sadržale disperzibilnu celulozu; tečni sorbitol; metil- i propil parahidroksibenzoat, glicerol i acesulfam-kalijum. Takođe su identifikovani i specifični ekscipijensi za PAR efervescentne tablete, tvrde kapsule, peroralne praškove, rastvore i suspenzije, kao i za IBU gelove i meke kapsule.

Ključne reči: *text mining*, farmaceutski oblici, kvalitativna analiza, ekscipijensi, paracetamol, ibuprofen