Journal Pre-proof

A topic model approach to identify and track emerging risks from beeswax adulteration in the media

Agnes Rortais, Federica Barrucci, Valeria Ercolano, Jens Linge, Anna Christodoulidou, Jean-Pierre Cravedi, Raquel Garcia Matas, Claude Saegerman, Lidija Svečnjak

PII: S0956-7135(20)30351-0

DOI: https://doi.org/10.1016/j.foodcont.2020.107435

Reference: JFCO 107435

To appear in: *Food Control*

Received Date: 1 April 2020

Revised Date: 9 June 2020

Accepted Date: 15 June 2020

Please cite this article as: Rortais A., Barrucci F., Ercolano V., Linge J., Christodoulidou A., Cravedi J.-P., Matas R.G., Saegerman C. & Svečnjak L., A topic model approach to identify and track emerging risks from beeswax adulteration in the media, *Food Control* (2020), doi: https://doi.org/10.1016/j.foodcont.2020.107435.

CRediT authorship contribution statement

Agnes Rortais: Conceptualization; Methodology; Validation; Investigation; Writing – original draft, review & editing ; Visualization; Supervision; Project administration. **Federica Barrucci**: Methodology; Writing - review & editing. **Valeria Ercolano**: Methodology; Data curation. **Jens Linge**: Data curation; Writing - review & editing**. Anna Christodoulidou**: Writing - review & editing**. Jean-Pierre Cravedi**: Writing - review & editing**. Raquel Garcia Matas**: Writing - review & editing**. Claude Saegerman**: Writing - review & editing. **Lidija Svečnjak**: Writing - review & editing**.**

Introduction

The European Food Safety Authority (EFSA) has the mission to establish procedures comprising tools and methodologies for emerging risks identification in food and feed (EC, 2002a). EFSA defines an emerging risk as “a risk resulting from a newly identified hazard to which a significant exposure may occur, or from an unexpected new or increased significant exposure and/or susceptibility to a known hazard” (EFSA, 2007), and an emerging issue as “an issue that has been very recently identified and merits further investigation to determine whether it meets the requirements of an emerging risk” (SCENIHR, 2009; EFSA, 2011). Linked to these definitions, i.e. whether the aim is to identify emerging emerging risks or issues, different tools and methodologies are available. Early-warning systems are used to monitor hazards deriving from a lack of compliance with existing regulations or to detect problems that are not yet regulated, but in some cases, they can be used to detect emerging risks when they are used to predict the development of the hazard (Marvin et al., 2009).

The Medical Information System from the Europe Media Monitor (EMM/MEDISYS), an automated system that monitors the media (Linge et al., 2009; Steinberger et al., 2013) was tested by EFSA as a tool to monitor emerging food- and feed-borne hazards (Rortais et al., 2010) and further developed for the monitoring and reporting of emerging plant pests such as the highly plant-pathogenic bacterium *Xylella fastidiosa* (Alomar et al., 2015, 2016; Ferilli et al., 2019). More recently, the system was developed to monitor food fraud media reports world-wide (MEDYSIS-FF) (Bouzembrak et al., 2018) that are published, via a monthly newsletter, by the Joint Research Centre. However, the analysis of the media in search of signals of emerging risks remains challenging given the amount of information and noise produced, but with the rise of artificial intelligence tools and modelling and Machine Learning Techniques (MLT), such challenges can be overcome (Ng et al., 2020). For example, topic models which can automatically explore large collections of documents, connect those that exhibit similar topics and deliver per topic associated patterns of words, generated with a probabilistic Latent Dirichlet Allocation (LDA) model (Blei & Lafferty, 2009; Blei, 2012), can support such searches.

Food frauds, which are of global concern and important drivers of food safety emerging risks, require to be better integrated into food safety assessments (EFSA, 2019). The term “food fraud” was formally defined by Spink & Moyer (2011) and later summarised as “an illegal deception for economic gain using food” (see Spink et al., 2019b for a review). Among the different types of identified food frauds, Spink et al. (2019a) found that adulterant-substances, defined as “any substances intentionally added to food, which are not present in such food as a result of the production”, were the most common. Recently and for the first time, EFSA received a mandate from the European Commission on a food fraud case, i.e. to assess the health concerns for both honey bees and humans to beeswax adulterated with paraffin and stearin (also referred as stearic and palmitic acids) and to define purity criteria for beeswax used in apiculture (EFSA, 2020).

Beeswax is a natural product made by honey bees to build comb wax and to store honey (called honeycomb), pollen and larvae (Bogdanov, 2016). In the apiculture sector, beeswax is considered as an animal by-product category 3 and, therefore, it is not intended for human consumption (EC, 2009). However, there are indications that such beeswax can enter the food supply chain, as seen with the selling of honeycombs (in pieces or inside honey pots) for consumption. This type of honey is authorised for sale on the market and known as "comb honey", "chunk honey" or "cut comb honey" (EC, 2002a, article 2). Therefore, with such practices in place and in the absence of an official analytical method to detect adulterants in beeswax, health safety risks to both honey bees and humans cannot be excluded.

This study aims at testing the use of LDA topic model to automatically classify large collection of results from searches made on the EMM/MEDISYS that are related to food fraud emerging risks and issues, taking beeswax adulteration as an example.

1. Material and methods

1.1. The EMM/MEDISYS

The EMM/MEDISYS is a near real-time news alert automated system managed by the Joint Research Centre of the European Commission. This system covers emerging and re-emerging public health issues related to communicable diseases and bioterrorism. Since its inception in 2004, the EMM/MEDISYS has extended its media coverage, improved information aggregation across documents and languages, and has gained a broad user base in the process. The EMM/MEDISYS enables users to moderate incoming news items and to disseminate the collected information. It monitors on a 24h/7d basis approximately 900 specialist medical sites plus all generic EMM news, i.e. over 20,000 RSS feeds and HTML pages sites from 7000 generic news portals and 20 commercial news wires in altogether 70 languages, including languages of the European Union (EU) Member States, Arabic, and Chinese.

1.2. Searches on the EMM/MEDISYS and filtering with LDA model

A retrospective search was made on the EMM/MEDYSIS system, in English, using the keyword "beeswax" from 01.01.2017 until 22.05.2019. A total of 3210 news articles were retrieved. Each news item contains meta data fields obtained from the original publisher such as link, title and description (i.e a shortened version of the full text limited to 300 characters). For the purpose of this study, access to the full text was obtained and processed.

After the removal of the duplicates (i.e. articles having the same title published by a different source), a total of 2276 news articles were screened using a LDA topic model. A graphical representation of the topic modelling algorithm is provided on Fig. 1. The topic modelling is an unsupervised MLT for organising, and summarizing massive collection of documents using probabilistic LDA model for uncovering the underlying semantic structure of a document collection. The LDA is based on two assumptions, first a collection of documents covers multiple topics and a document contains one or more topics. The aim of topic modelling algorithm is to use the original text documents to infer the hidden topic structure that likely generate the observed collection of documents. Practically a topic is defined as a distribution of frequencies over a fixed vocabulary ($\varphi(k)$, see Fig. 1) and each word in each document ($W(i,j)$ see Fig. 1) is drawn from one of the vocabularies of the topics appearing in the document, (where the topic presence follows the $\Theta(i)$ distribution, see Fig. 1).

Since no assumption about the underline structure of the corpus was available, the LDA model was chosen, assuming independence in the occurrence of topics, i.e. the probability that a topic is present in a document is not affected by the presence in the document of any other topics (Blei and Lafferty, 2009).

Each document in the collection was pre-processed removing stop words, word with lenght smaller than 3 letters, numbers and punctuation and reducing words to their stem. Then a document-term matrix, a matrix that describes the frequency of terms that occur in a collection of documents, was generated and used as input for LDA model. Only terms which have a tf-idf (term frequency-inverse document frequency, a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus) value bigger than the median were included. R package topic models (Grün and Hornik, 2011) was used to fit LDA topic model to the document-term matrix.

Optimal number of topics was set based on best performance of 4 indexes proposed by Griffiths and Steyvers (2004), Arun et al. (2010), Deveaud et al. (2014) and Juan et al. (2020). The optimal number of topics ranged between 10 and 35. Like others unsupervised MLT, the choice of the optimal number is based on validation, consisting in checking if topics are well defined and separated. The best resulting choice was 10 topics. Furthermore, to determine how the topics were related to adulterated beeswax and effects on bee and human health, the patterns of words with their frequencies associated to each topic were checked for the presence of the words "health ", "beeswax", "wax", "food", "bee" and "beekeeper". Depending on the number of matching words displayed in each topic, six levels of relatedness were determined, i.e. "not relevant" when no word matched and "very low", "low", "medium", "high" or "very high" when one, two, three, four or five words, respectively, matched among the 30 most frequent words generated per topic. Finally, to validate topic relatedness to adulterated beeswax, one expert screened the 2276 news articles, first by titles and description, and, second (if unclear or deemed relevant), by reading the full text article.

2. Results and discussion

2.1. Identification of beeswax topics in the media

The fitting of the LDA topic modelling resulted in the identification of 10 topics (Fig. 2.A-F) with a total of 2276 news articles distributed across those 10 topics (Fig. 3), showing various levels of relatedness to beeswax adulteration and effects on bee and human health (Table 1).

The three topics that were not found related to beeswax adulteration (i.e. topics 1, 4 and 10 in Table 1) were picked up with the search keyword "beeswax" because they were related to other uses of the word "beeswax" such as colloquial (e.g. "mind your beeswax") and common names (e.g. "Beeswax ship", "Beeswax Dyson Farming", "beeswax advertising startup", "Ms Beeswax", "Beeswax Park", etc.).

The two topics that were found the most ("highly" and "very highly") related to beeswax adulteration (i.e. topics 5 and 8 in Table 1) contained indeed relevant information. Topic 5 contained all official reports from the European Commission (EC, 2019), the European Parliament (EP, 2017a, b, 2018a, b, c, d) and the Food Safety and Standards Authority of India (FSSAI, 2018) reporting on the issue of beeswax adulteration by stearin and paraffin in Europe. Topic 8 contained news articles on adulteration in beeswax (Zhou, 2018).

The five remaining topics that were found less ("low" and "very low") related to beeswax adulteration (i.e. topics 2, 3, 6, 7 and 9 in Table 1) contained news articles related to beeswax market and trends (emerging beeswax market opportunities ; topic 9) and to the use of beeswax as food packaging (beeswax wrapping paper ; topic 2), food additive (beeswax as wine additive ; topic 3) and in other non food-related sectors (e.g. beeswax in cosmetics in topic 7 and beeswax used as encaustic and mould for sculptures in topic 6) (Table 2).

Finally, within the 10 topics, no article was found on effects of beeswax adulteration on bee health or human health.

2.2. Topic model to reduce noise in the media on beeswax adulteration

The identification of several reports on beeswax adulteration from official sources (i.e. EP, EC and the FSSAI demonstrates that the EMM/MEDYSIS is able to monitor food fraud related to beeswax adulteration. To monitor automatically and continuously beeswax adulteration on the EMM/MEDYSIS, a specific filter needs to be developed, which is time consuming. Topic modelling allowed to identify information that could be used for the improvement of such a filter, i.e. to reduce noise. Further optimization of such a filter on the EMM/MEDYSIS could include a broader coverage in terms of languages (up to 70 languages) and sources (e.g. websites related to the apiculture sector). This filter could be added to the list of Food Fraud (FF) reports published world-wide in the media (MEDYSIS-FF; Bouzembrak et al., 2018).

2.3. Topic model to reduce information to be assessed in the media on beeswax adulteration

The results of this study showed that topic modelling allowed to explore, organise and summarise rapidly and efficiently a high number of news articles and reports from the media on beeswax adulteration. Indeed, topic modelling allowed to filter out 13.8% of the news articles from the original retreived articles (i.e. all articles from the topics that were not found related to beeswax adulteration). When considering only the two topics that were the most related to beeswax adulteration, topic modelling allowed to filter out 77.5% of the articles. For the setting of a filter to monitor fraud related to beeswax adulteration, this method proved extremely useful with the rapid identification of both irrelevant (noise) and relevant information (official reports and news articles).

2.4. Topic model to identify emerging risks in the media related to beeswax adulteration

The news article on the adulteration of beeswax (Zhou, 2018) provided a signal of an emerging risk corresponding to "a risk from an unexpected new or increased significant exposure" to beeswax adulterants for humans and bees. Indeed, Zhou (2018) alerted that the beeswax did not comply with the expected quality standards […"I need to get some of that beeswax tested," Rice said. "That wax was supposedly Australian, filtered, cosmetic grade. I've spoken to other people at the bee club that say that's not right."]. Further, the author explained that adulterated beeswax may enter the beeswax supply chain because beeswax price is continously increasing in response to a growing demand for beeswax from the cosmetic and food packaging (wrapping paper) sectors […New uses for the wax – from cosmetics to food wraps – and the comparative health of Australia's bees have driven the export price of Australian beeswax up in the global marketplace …"It suggests to me that this is probably being sold because the price is so attractive. People are trying to scam it and trying to make big bucks."]. These trends were confirmed when screening the other articles under the topics related to the use of eco-friendly beeswax wrapping paper to replace plastic use (Byrnes, 2019), the use of natural cosmetics made of beeswax (Denham, 2020), and the global beeswax market forescast 2018-2025 (Market Watch, 2018). Finally, another factor that contributes to the increase in beeswax value and price is the problem of the decline of bee populations worldwide.

A further analysis of the articles dealing with the use of beeswax in cosmetics and food packaging, highlighted other trends which could be new emerging risk pathways linked to adulterated beeswax. The beeswax market in cosmetics is evolving fast, promoting high-quality products at low costs, imposing significant changes in the sector. As a result, key emerging markets and new opportunities in the area of beeswax are developing, as seen with the production of cosmetic kits and other beeswax products (e.g. food wrappers), all available online. Also, in the area of cosmetics, several news articles promoting home-made recipies distributed on internet were found. Those practices could pose a safety concerns to consumers if the origin, quality and safety of the beeswax is not controlled and cannot be guaranteed.

From the media and the scientific literature, food fraud incidents are usually linked to gaps in quality assurance testing methodologies and inadequate existing regulatory systems (Spink and Moyer, 2011). Currently, there is no legislation on beeswax intended for use in apiculture and as food (honeycomb) as the beeswax is considered as an animal-by-product category 3 and therefore not for human consumption. Potential risks to both human and bee health linked to adulterated beeswax entering the apiculture sector was assessed by EFSA (2020) and showed the need for a revision of the current legislation, for additional data to fill the gaps and to conduct more evidence-based risk assessment in this area. Regarding the use of beeswax wrapping paper, if the beeswax is adulterated and adulterants migrate to the food in contact with the wrapping paper, this might also present a safety concern for humans. In the case of cosmetics, although this is not a food-related exposure and it is covered by legislation (EC, 2012), a more careful analysis of this new potential pathway could provide new information (hazards and processes) on beeswax adulteration. In sectors where beeswax is being subjected to routine quality (purity) testing (pharmaceuticals, cosmetics and food industry for beeswax used as food additive), there is also a risk of not detecting adulterants (e.g. paraffin, stearin/stearic acid, and carnauba wax) in beeswax due to detection limits, imposed by standardized physico-chemical analytical methods, that are too high (i.e. between 5-50%) (Svečnjak et al., 2019).

3. Conclusions and perspectives

In conclusion, topic modelling proved efficient in reducing noise and information from the media to be thereafter more rapidly processed by human intervention. It provided support for the definition of a more specific filter on beeswax adulteration in

EMM/MEDISYS, that could be added to the MEDISYS-FF list, and it proved useful in identifying emerging risks linked to beeswax adulteration, i.e. through new exposure pathways and possibly through new hazards if further analysis is conducted beyond the apiculture sector. The increase of languages and sources' coverage should also refine the model being more accurate in terms of scale and space-time occurence of adulteration of beeswax.

Although this study needs further fine tuning (e.g. implementing different topic models algorithm in addition to LDA, to relax the assumption made such as independance between topics), those preliminary results demonstrate the usefulness of topic modelling in identifying rapidly topics relevant to food fraud incidents in the media corpus. The application of the topic model MLT on the articles retrieved through the MEDYSIS-FF, e.g. through an annotated algorithm developed in R package (R Core Team, 2017), should permit to quickly identify, alert on, and track emerging risks from beeswax adulteration. This information could be used (e.g. on a monthly basis) by risk assessors, decision-makers, representatives of the beekeeping sector and scientists to better detect and trace food frauds in the food supply chain as well as to understand the underlying mechanisms and how to mitigate their impacts.

References

Alomar, O., Batlle, A., Brunett, J. M., García, R., Gil, R., Granollers, T., Jiménez, S., Laviña, A., Linge, J. P., Pautasso, M., Reverté, C., Riudavets, J., Rortais, A., Stancanelli, G., Volano S., & Vos, S. (2015). Development and testing of the media monitoring tool MedISys for early identification and reporting of existing and emerging plant health threats. *EPPO Bulletin*, 45, 288-293. doi.org/10.1111/epp.12209

Alomar, O., Batlle, A., Brunett, J. M., García, R., Gil, R., Granollers, T., Jiménez, S., Laviña, A., Reverté, C., Riudavets, J., & Virgili-Goma, J. (2016). Development and testing of the media monitoring tool MedISys for the monitoring, early identification and reporting of existing and emerging plant health threats. EFSA Supporting Publications 2016:*13*(12):EN-1118. doi.org/10.2903/sp.efsa.2016.EN-1118

Arun, R., Suresh, V., Madhavan, C.V., & Murthy M.N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 391-402). Berlin Heidelberg: Springer. http://doi.org/10.1007/978-3-642-13657-3_43

Bernal, J. L., Jiménez, J. J., Del Nozal, M. J., Toribio, L., Martín, M. T. (2005). Physico-chemical parameters for the characterization of pure beeswax and detection of adulterations. European Journal of Lipid Science and Technology 107(3): 158-166. Blei, D. M., & Lafferty, J. D. (2006). Topic Models. In A. Srivastava & M. Sahami (Eds). Text mining: classification, clustering, and applications. London: Chapman & Hall/CRC Data Mining and Knowledge Discovery Series.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM,* 55(4), 77-84.

Blei, D.M., & Lafferty, J.D. (2009). Topic Models. In A Srivastava & M Sahami (eds.), *Text Mining: Classification, Clustering, and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series.

Bogdanov, S. (2016). Beeswax: History, uses and trade. Beeswax book (Chapter 2). Bee Product Science, pp. 1–17.

Bouzembrak, Y., Steen, B., Nelso, R., Linge, J., Mojtahed, V., & Marvin, H. J. P. (2018). Development of food fraud media monitoring system based on text mining. *Food Control*, *93*, 283-296.

Byrnes, H. (2019, April 19). Genius inventions that can stop us from using so much platsic. *MSN News*, retrieved from https://www.msn.com

Cheong, J. I. F. (2017, September 29). The Patriarch of Germany II. *MacauDaily Times*, retrieved from https://macaudailytimes.com.mo/patriarch-germany-ii.html

Denham, J. (2020, March 9). 15 best lip balms. *Independant*, retrieved from https://www.independent.co.uk

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique, Lavoisier*, 17(1), 61–84. http://doi.org/10.3166/dn.17.1.61-84

European Commission (EC) (2002a). Regulation (EC) No. 178/2002 of the European Parliament and of the Council. Official Journal of the European Communities, 1.2.2002 L 31/1.

European Commission (EC) (2002b). Council Directive 2001/110/EC. Official Journal of the European Communities, 12.1.2002 L10 47.
European Commission (EC) (2009). Annex XIV, Table 2 of Regulation (EC) No 1069/2009 of the European Parliament and of the Council of 21 October 2009 laying down health rules as regards animal by-products and derived products not intended for human consumption and repealing Regulation (EC) No 1774/2002 (Animal by-products Regulation) OJ L 300, 14.11.2009, p. 1–33
European Commission (EC) (2012). Commission Regulation (EU) No 231/2012 of 9 March 2012 laying down specifications for food additives listed in Annexes II and III to Regulation (EC) No 1333/2008 of the European Parliament and of the Council Text with EEA relevance OJ L 83, 22.3.2012, p. 1–295. ELI: http://data.europa.eu/eli/reg/2012/231/oj
European Commission (2019). Honey – Beekeeping. EC publication on 17.04.2019 at https://ec.europa.eu/info/food-farming-fisheries/animals-and-animal-products/animal-products/honey_en
European Food Safety Authority (EFSA) (2007). Definition and description of “emerging risks” within the ESFA’s mandate. Scientific Committe and Advisory Forum Unit, Parma, 10 July 2007, 3pp.
European Food Safety Authority (EFSA) (2011). EFSA's 15th Scientific Colloquium on emerging risks in food: from identification to communication. EFSA Scientific Colloquium Summary Report, 12-13 October 2010, Parma, Italy. https://doi.org/10.2903/sp.efsa.2011.EN-114
European Food Safety Authority (EFSA), Afonso, A., Garcia Matas, R., Maggiore, A., Merten, C., Yin, A., Robinson, T. (2019). Technical report on EFSA’s activities on emerging risks in 2018 EFSA supporting publication 2019:EN-1704, 49 pp. doi:10.2903/sp.efsa.2019.EN-1704
European Food Safety Authority (EFSA) (2020). Risk assessment of beeswax adulterated with paraffin and/or stearin/stearic acid when used in apiculture and as food (honeycomb). EFSA Supporting Publication 2020:EN-1859. 64pp. DOI:10.2903/sp.efsa.2020.EN-1859
European Parliamant (2017a). Written question - Protection of apiculture products other than honey under EU quality schemes - E-003443/2017. EP publication pn 17.05.2017 at https://www.europarl.europa.eu/doceo/document/E-8-2017-003443_EN.html?redirect
European Parliamant (2017b). Answer - Protection of apiculture products other than honey under EU quality schemes - E-003443/2017. EP publication on 11.07.2017 at http://www.europarl.europa.eu/sides/getAllAnswers.do?reference=E-2017-003443&language=EN
European Parliamant (2018a). Report - Prospects and challenges for the EU apiculture sector - A8-0014/2018 - Committee on Agriculture and Rural Development. EP publication on 08.02.2018 at https://www.europarl.europa.eu/doceo/document/A-8-2018-0014_EN.html?redirect
European Parliamant (2018b). Written question - Stearin in beeswax - E-002661/2018. EP publication on 17.05.2018 at http://www.europarl.europa.eu/sides/getDoc.do?type=WQ&reference=E-2018-002661&format=XML&language=EN

European Parliamant (2018c). Answer - Stearin in beeswax - E-002661/2018. EP publication on 06.08.2018 at http://www.europarl.europa.eu/sides/getAllAnswers.do?reference=E-2018-002661&language=EN

European Parliamant (2018d). Notice to members - Petition No 0432/2016 by Zoltán Tóth (Hungarian) on stopping the commercial sale of honey products manipulated by macro-reticulated synthetic resin, which endanger the health of European consumers - PE 605.989v02-00 - Committee on Petitions. EP publication on 30.11.2018 at http://www.europarl.europa.eu/sides/getDoc.do?type=COMPARL&reference=PE-605.989&format=PDF&language=EN&secondRef=02

Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. *Journal of Food Protection, 76*(4), 723-35. doi: 10.4315/0362-028X.JFP-12-399.

Federal Agency for the Safety of the Food Chain (FASFC) (2018). Advice 18 2018 of the Scientific Committee of the FASFC regarding the risk to bee health of contaminated and adulterated beeswax. http://www.afsca.be/comitescientifique/avis/2018/_documents/Avis18-2018_SciCom2016-27_residus_cire_santeabeilles.pdf

Ferilli, F., Stancanelli, G., Ling, J. P. & Mannino, M. R. (2019). A new online resource to monitor new or emerging plant pests: MedISys media monitoring and the case of *Xylella fastidiosa*. *Phytopathology*, *109*(2), 216-218. https://doi.10.1094/PHYTO-07-18-0241-A.

FSSAI (2018, August 22). FSSAI notifies food safety standards for honey and its products to curb adulteration. *Firstpost*. Retrieved from https://www.firstpost.com/business/fssai-notifies-food-safety-standards-for-honey-and-its-products-to-curb-adulteration-5019571.html

Glowatz, E. (2017, April 25). Caterpillars could stop plastic bag pollution. *International Business Times*, retrieved from http://www.ibtimes.com

Griffiths, T.L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101 (suppl 1), 5228–5235. http://doi.org/10.1073/pnas.0307752101

Grün B, Hornik K (2011). Topic models: An R Package for Fitting Topic Models. *Journal of Statistical Software*, **40**(13), 1–30. doi: 10.18637/jss.v040.i13.

Juan, C., Tian, X., Li, J., Zhang Y., & Tang S. (2009). A density-based method for adaptive lda model selection. *Neurocomputing — 16th European Symposium on Artificial Neural Networks 2008*, 72(7–9), 1775–1781. http://doi.org/10.1016/j.neucom.2008.06.011

Kameda, T. (2004). Molecular structure of crude beeswax studied by solid-state 13C NMR. *Journal of Insect Science* 4(29), 1-5. https://doi.org/10.1673/031.004.2901

Kocaeli (2018, March 2). Ottoman painter's house becomes museum. *Hurriyet Daily News*, retrieved from http://www.hurriyetdailynews.com

Linge, J. P., Steinberger, R., Weber, T. P., Yangarber, R., van der Goot, E., Al Khudhairy, D. H., & Stilianakis, N. (2009). Internet surveillance systems for early alerting of health threats. *Eurosurveillance*, *14*, 19162.

Market Watch. Beeswax - Global Market Outlook 2018-2025: Growth Opportunities, Competition, Trends & Outlook Across Applications & Regions - ResearchAndMarkets.com. Market Watch (2018). https://www.marketwatch.com/press-release/beeswax---global-market-outlook-2018-2025-growth-opportunities-competition-trends-outlook-across-applications-regions---researchandmarketscom-2018-10-18. Accessed 18 October 2018.
Marvin, H.J.P, Kleter, G.A., Prandini, A., Dekkers, S., & Bolton, D.J., (2009). Early identification systems for emerging foodborne hazards. *Food and Chemical Toxicology*, 47, 915-926.
Navarro-Hortal, M. D., Orantes-Bermejo, F. J., Sanchez-Gonzales, C., Varela-Lopez, A., Giampieri, F., Fernandez-Pinar, C. T., et al. (2019). Industrial-Scale Decontamination Procedure Effects on the Content of Acaricides, Heavy Metals and Antioxidant Capacity of Beeswax. *Molecules*, *24*(8), 1518. https://doi.org/10.3390/molecules24081518
Ng, V., Rees, E.E., Niu, J., Zaghlool, A., Ghiasbeglou, H., Verster, A. (2020). Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. *Canada Communicable Disease Report*, *46*(6):186–91. https://doi.org/10.14745/ccdr.v46i06a06
R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Rortais, A., Belyaeva, J., Gemo, M., van der Goot, E., & Linge, J. (2010). MedISys: An early-warning system for the detection of (re-)emerging food- and feed-borne hazards, 2010. *Food Research International*, 43(5), 1553-1556. https://doi: 10.1016/j.foodres.2010.04.009.
Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR), 2009. Emerging Issues and the Role of the SCENIHR. Position Paper.
Spink J., & Moyer D.C., (2011). Defining the public health threat of food fraud. J. Food Sci. 2011;76:R157–R162. doi: 10.1111/j.1750-3841.2011.02417.x.
Spink, J., Embarek, P. B., Savelli, C. J., & Bradshaw, A. (2019a). Global perspectives on food fraud: results from a WHO survey of members of the International Food Safety Authorities Network (INFOSAN). NPJ science of food, 3, 12. https://doi.org/10.1038/s41538-019-0044-x
Spink, J., Bedard, B., Keogh, J., Moyer, D.C., Scimeca, J. and Vasan, A. (2019b). International Survey of Food Fraud and Related Terminology: Preliminary Results and Discussion. Food Science. 84(10), 2705-2718. https://doi.org/10.1111/1750-3841.14705
Steinberger, R., Pouliquen, B., & van der Goot, E. (2013). An introduction to the Europe Media Monitor family of applications. https://doi:1309.5290.
Svečnjak, L., Chesson, L.A., Gallina, A., Maia, M.. Martinello, M., Mutinelli, F., Necati Muz, M., Nunes, F.M., Saucy, F, Tipple, B.J., Wallner, K., Waś E., & Waters, T.A., (2019). Standard methods for Apis mellifera beeswax research. In Dietemann, V., Neumann, P., Carreck, N., Ellis, J.D. (Ur.). The COLOSS BEEBOOK. Volume III. Part I: Standard methods for Apis mellifera hive products research. Journal of apicultural research. 58 (2): 1-108.
University of California - Los Angeles (2017, December 11). Scientists pioneer new way to analyze ancient artwork: "Macroscale multimodal chemical imaging' reveals

details about second century Egyptian painting." *ScienceDaily*, retreived from www.sciencedaily.com/releases/2017/12/171211145912.htm

Zhou, N. (2018, May 31). Mind your beeswax: global price surge leaves bearded Australians in a tangle. *The Guardian*. Retrieved from https://www.theguardian.com/environment/2018/may/31/mind-your-beeswax-global-price-surge-leaves-australian-beard-owners-in-a-tangle

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared ro influence the work reported in this paper.

Acknowledgement
This research did not receive any specific grant from funding agencies in the public, commercial, or not-profit sectors

Disclaimer
The views or positions expressed in this article do not necessarily represent in legal terms the official position of the European Food Safety Authority (EFSA). EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear. This article does not disclose any confidential information or data. Mention of proprietary products is solely for the purpose of providing specific information and does not constitute an endorsement or a recommendation by EFSA for their use.

List of captions

Fig. 1. Graphical representation of Topic Modelling - LDA - (©Naskar (2019, February 14). Latent Dirichlet Allocation for Beginners: A high level overview. *NLP*. Retrieved from https://www.thinkinfi.com/2019/02/lda-theory.html).

Fig. 2. Patterns of words for the 10 identified topics.

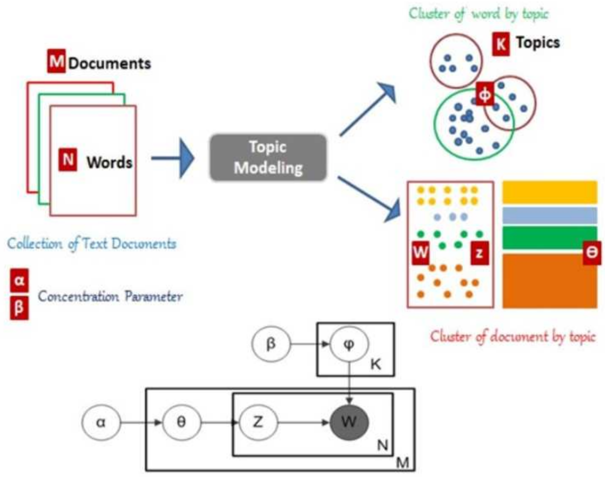
Fig. 3. Number of articles retrieved per topic (1 to 10) on EMM/MEDISYS with the keyword "beeswax".

Table 1. Topics relatedness to beeswax adulteration according to the number of words present (scored 1) or absent (scored 0) in the 10 topics.

Table 2. Main issues described in the news articles contained in the 7 topics related to beeswax.

Topics	Words in topics						Relatedness (number of words)
	Health	Beeswax	Wax	Food	Bee	Beekeeper	
1	0	0	0	0	0	0	Not relevant (0)
2	0	1	0	1	0	0	Low (2)
3	0	1	1	0	0	0	Low (2)
4	0	0	0	0	0	0	Not relevant (0)
5	1	1	0	1	1	1	Very high (5)
6	0	1	0	0	0	0	Very low (1)
7	0	1	0	0	0	0	Very low (1)
8	0	1	0	1	1	1	High (4)
9	0	1	0	0	0	0	Very low (1)
10	0	0	0	0	0	0	Not relevant (0)

Topic	Main issues
2	Eco-friendly beeswax wrap to reduce plastic use (Byrnes, 2019)
3	Caterpillar eating plastic (Glowatz, 2017), beeswax wine additive (Cheong, 2017), beeswax chemical composition (Kameda, 2004)
5	Official report on beeswax adulteration from EP (2017a, b, 2018a, b, c, d), EC (2019) & FSSAI (2018) and research on beeswax analytics (Navarro-Hortal et al., 2019)
6	Beeswax used as encaustic in art painting (University of California – Los Angeles, 2017), beeswax sculptures in museum (Kocaeli, 2018)
7	Natural, organic and vegan cosmetics (lip balm, cream) (Denham, 2020)
8	Adulteration in beeswax and cosmetics (Zhou, 2018)
9	Global beeswax market forecast 2018-2025 and emerging beeswax market opportunities (Market Watch, 2018)



The LDA algorithm inputs are **M** no. of documents, each of these documents have **N** no. of words. Then two parameters are fixed, respectively:

α : Per-document topic distribution (high α means every document is likely to contain a mixture of most of the topics, low α means that a document is likely to be represented by just a few topics);

β : Per topic word distribution (high β means that each topic is likely to contain a mixture of most of the words, low β means topics may contain a mixture of just a few of words)

The algorithm processes these input parameters in an iterative way realizing two main actions:

1) Cluster of words by topic:

K: no. of topics

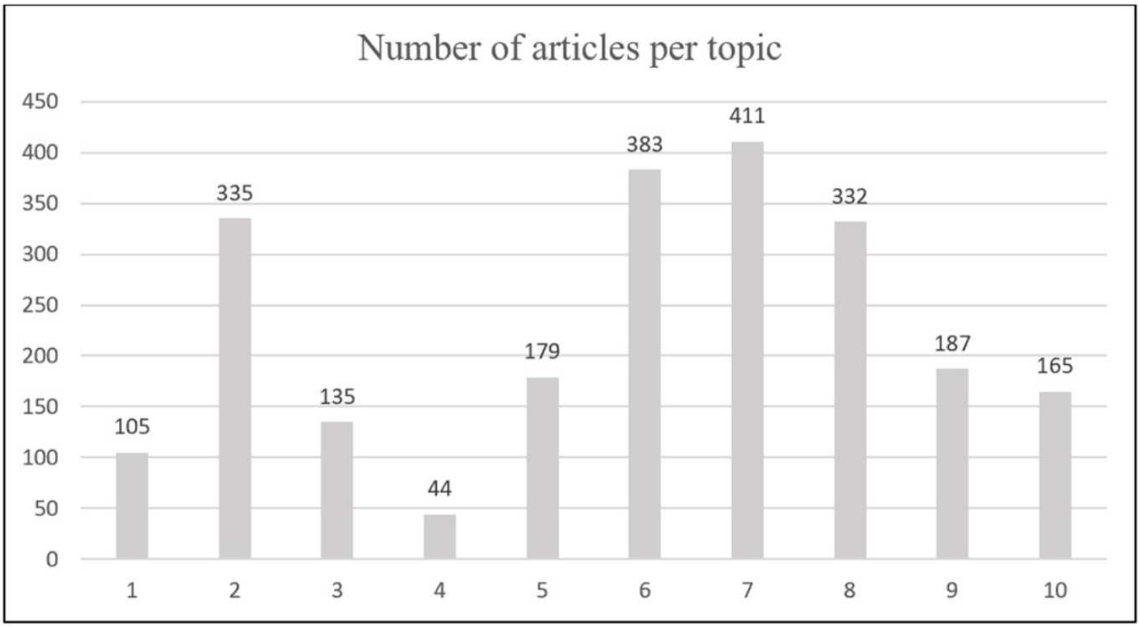
ϕ(k): w distribution for topic k (topic probability per word)

2) Cluster of documents by topic:

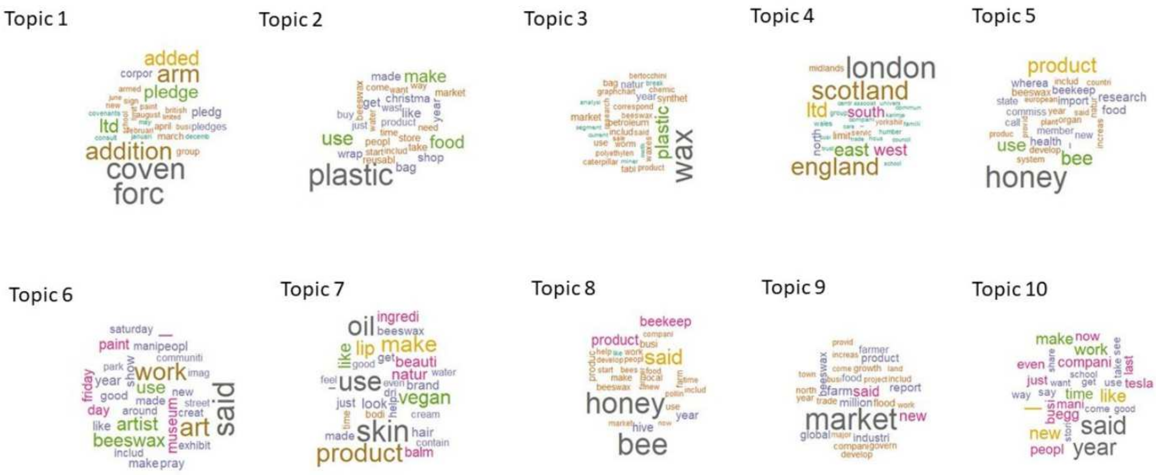
W = W(i,j): j-th word in i-th document

Z = Z(i,j): topic assignment for w(i,j) word

Θ(i): topic distribution for document i (topic probability per document)



Number of articles per topic
450
400
350
300
250
200
150
100
50
0
105
335
135
44
179
383
411
332
187
165
1
2
3
4
5
6
7
8
9
10


Topic 1
added
arm
pledge
ltd
addition
coven
forc
Topic 2
make
use
food
plastic
Topic 3
plastic
wax
Topic 4
london
scotland
ltd
south
east
west
england
Topic 5
product
use
bee
honey
Topic 6
work
artist
art
beeswax
said
Topic 7
oil
make
lip
use
vegan
skin
product
Topic 8
beekeep
product
said
honey
bee
Topic 9
market
Topic 10
make
work
companI
new
said
year

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: