Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2022) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)

Expanding Export Control-related Data for Expert System

Akihiko Obayashi[a*] and Rafal Rzepka[b]

[a]*Center for Innovation and Business Promotion, Hokkaido University, Kita-ku, North 21, West 10, Sapporo, Japan*
[b]*Faculty of Information Science and Technology, Hokkaido University, Kita-ku, North 14, West 8, Sapporo, Japan*

Abstract

In this paper we describe an extension of question answering dataset to be utilized for testing security export control expert system. Differently to the long questions and answers with rich context mostly describing exceptional interpretation of regulatory texts, our addition contains short questions and answers which were manually created by an expert who often answers similar questions from academic researchers. This dataset extension should allow broader and more realistic experiments for evaluating dialog-based expert systems with more realistic input. The dataset itself could be used broadly in other studies as a testbed for testing both text classification methods and QA systems requiring external data processing. We introduce the task, problems of previously used data, and describe how they are tackled with the additional set. Examples are presented together with term frequency analysis. Unlike the government-made QAs, the newly created dataset is freely available upon request.

Keywords: expert system; export control; question answering; data creation;

1. Introduction

Recent two decades have shown an increasing necessity for Security Export Control. Since the Wassenaar Arrangement established in 1996, controlling transfer of technologies and export of goods for the purpose of preserving the peace and security of the international community became a standard in the Western world. The importance of this field lays in preventing transfer of the technologies and goods that could be used for military purposes. Avoiding transferring items to parties who could use them for terrorist attacks or threaten the peace and security of the international community becomes crucial not only for business but also for the academic world. In the United States, universities often resort to DoResearch, online decision tree interface helping academics to decide if their work could be a potential threat and if there is a need to apply for an export permission. However, this system requires users to have specialized knowledge of export control, which makes it perplexing to those who lack extensive training in the field. The goal of our research is to build a dialog system in which experts in export control and artificial intelligence collaborate to develop a novel user-friendly support for non-experts. For testing our system under development, we have extracted question-answer pairs from existing governmental documents. However, the data, although rich in information, does not fully mirror the whole spectrum of possible questions asked to the system by potential users. The questions are sent to the specialists who write elaborate answers. Because it is less likely scenario for a conversational exchange of information, we added 87 new pairs to extend the dataset capability for testing the QA system and experiment with answer retrieval methods. Section 2 of this paper introduces related research; Section 3 describes the data preparation process and

Section 4 concentrates on performed annotation including detailed explanation and examples which were lacking in the previous publication. Section 5 concludes the paper and introduces our plans for the future work.

2. Related Work

In the advent of deep learning, neural language models like BERT [2], RoBERTa [3] or GPT-3 [4] have caused a shift from narrow task-oriented question answering tasks to large, most often automatically or half-automatically created datasets. They tend to be based on Wikipedia or crowdsourcing, which do not require any expertise in both creation and evaluation of the utilized data. Even when it comes to specific topics, the approach remains broad and automatic. For example in QA system for European Union regulations [5], the existing labels are used - “Causal Agent” (employee, consumer, etc.), “Organization” or “Time Period”, etc. and answers are limited to pointing to the document ID numbers. In the past, more specific questions and answers, often in shape of FAQs (Frequently Asked Questions), were more frequently developed. Nevertheless, nowadays simpler tasks seem to be preferred, also for clearer benchmarking and comparison of algorithms. One of such examples is limiting the dataset to recognizing question entailment [6]. One of the main reasons for this shift in the task for retrieving answers from texts is the fact that expert annotations are required [7]. This process is costly, especially with the amounts required by the data-hungry machine learning approaches. In some cases, crowdworkers can retrieve a correct answer to e.g., what-, how-, why-questions, when provided with the part of the document which contains the answer [8]. In other cases, annotators can be fully automated like in UIMA pipeline [9] or Appi service [10]. However, when the answer is not obvious and requires multi-hop reasoning with referring to the different parts of the regulatory documents, the level of difficulty increases, and the number of such datasets drastically decreases. One possible reason for this phenomenon is that the current methods are probably not yet capable of such complicated processing. Our work aims at providing a novel QA dataset for Japanese language, which can be widely used also for testing other algorithms.

3. Data Preparation

In this section we describe the original QA dataset and its expansion.

3.1. CISTEC Guidelines QA Dataset

The dataset introduced in our previous work [1] is based on guidelines for security export control published by the Center for Information on Security Trade Control (CISTEC). Its target audience is people involved in security export control at companies, universities, etc. The booklet provides detailed information required to determine whether a shipment should be controlled or not. The guidance assumes that audience have some basic knowledge of security export control. We manually extracted the QA data which is provided within the guidance pdf files omitting these with pictures and ones referring to other documents without giving direct answers. In total we have extracted 548 question-answer pairs. However, the questions are very detailed, and the answers go far beyond the contents of the legal texts we use as the base knowledge in our system. Therefore, we decided to extend this set with a subset containing shorter questions and answers for further experiments. In our previous work [2], we tested the CISTEC QA data with various machine learning methods to retrieve and generate answers, but because shorter inputs are also very probable and they cause new problems (shortage of contextual clues, for example), we created our own QA pairs described in the next subsection.

3.2. In House QA Dataset

There is one more problem with the CISTEC QA data explained above. In fact, it does not necessarily serve the purpose of providing support limited to the relevant classification. In other words, while the relevant classification is related to the list control[1], it also includes those related to the catch-all control[2]. To give a specific example, the description of the toluene regulation falls into the catch-all category. The description related to the catch-all regulation is considered to be information that is close to noise for the purpose of our system to support the determination of list control. Therefore, we have decided to add typical question-answer pairs that are focused on the relevant list control and the classification is performed by an expert (the first author of this paper). In this sense, it is obvious that the added pairs have a high ratio of being either "Applicable" or "Not applicable”. For example, in response to the question, "What kind of supercomputers are controlled?”, the answer is “In general, high-performance ones are controlled. Specifically, it is defined as those exceeding 29 weighted tera FLOPS." The answer is straightforward, and the label is "Applicable”, meaning that the asker should apply for a permission to export such supercomputer or to share detailed information about its technical details. In response to the question, "What types of resistors are regulated?”. The answer to the question is also

[1] Controlled items based on international export regimes and all countries/regions.
[2] All items other than controlled items and countries/regions other than Group A.

clear: "Neither fixed nor variable resistors are regulated." (and the label is "Not applicable"). On the other hand, to the question, "What types of robots are regulated?" The answer is "Robots are regulated by the Export Ordinance, including Item 2, Item 6, Item 12, etc. In general, explosion-proof and radiation-proof robots are controlled." Apparently, this answer provides important information to those who want to know the classification of robots. However, whether the label can be regarded as "Applicable" or not may require more consideration or further (and more detailed) investigation.

In total, 87 new question-answer pairs were created, and the annotation was performed, which is explained in the next section. More examples with their translations to English are given in Appendix A.

4. Expert Annotation

In this section we present details of the annotation process.

4.1. Annotation Labels

The labels used are the same as those used in our previous work [1] for the answer part of the CISTEC QA data. The labels are annotated into controlled ("Applicable"), not-controlled ("Not applicable"), "Requiring confirmation", and miscellaneous ("Other") depending on the answer content. The label "Requiring confirmation" is for answers for which the details are to be determined from the context of the ministerial ordinance. In the dialog-based expert system scenario, the chatbot can try to generate further inquiries to determine more details or ask the user to contact the human expert.

In addition, we have added the labels for the simpler questions added by the expert. Each question was annotated with one of two labels: "Request for explanation" or "Yes/No". In the example explained in 3.2, supercomputers and resistors are divided into "Applicable" and "Not applicable" respectively as the label of answer part, but the question type for both cases is identical: "Request for explanation". However, in case of "Is the supercomputer controlled?", the question type becomes "Yes/No". The effectiveness of this classification is a part of our future work.

4.2. Annotation Strategy

In 3.2, we explained that the expert added typical QA pairs focused on the relevant classification which is based on his career experience and questions he was asked in the past. Therefore, it is reasonable that a high percentage of the added QA are labelled either "Applicable" or "Not applicable". In this sense, they are regular question-answer pairs. On the other hand, the expert has also added a question "What do you mean by 'designed for space'?" Corresponding answer is "a product that is designed or manufactured to operate at an altitude of more than 100 kilometers above the earth's surface, or that the product has passed a test and is qualified". This pair was added because it contains important information for the classification relevant in the field of space technology. In other words, "designed for space" requires the threshold of an altitude of over 100 kilometers. Thus, the answer cannot be explicitly labeled neither "Controlled" nor "Not controlled", and the "Other" label is used instead.

For the proposed dataset, we have added a new label "No permission required" which means that an item is applicable but as a special exception does not require permission from the Minister of Economy, Trade and Industry.

4.3. Data-specific difficulties

Creating a dataset for e.g. common sense question answering task with crowdworkers is obviously effective for a standard QA system, but it is problematic in our case as it requires quite rare expertise. This is especially visible in our case of the trade security in academic environment as Japanese university began almost 30 years after the companies started to manage security export control (2016 and 1987, respectively). Furthermore, preparing an abundant number of, for example, computer-related QA pairs (or using existing FAQs) alone may not be useful for our expert dialogue system which traverse multiple fields and touches very specific aspects of related knowledge. It is necessary to prepare QAs for a wide range of fields, from electronics to biotechnology, nuclear power, and space engineering, which encompass security export control as a whole.

To save the expert labor, it may be possible (at least in part) to automatically generate typical QA utilizing regulatory texts and regulations. Documents containing laws and regulations basically inform what is regulated, and exemptions stipulate what is not regulated. In other words, the texts in principle define what is controlled. This could help with simplifying the QA task, but in real life, the assumption that what is not mentioned is not regulated might be pernicious. This difficulty, also related to the problem of responsibility, led us to create the dataset which is demanding for current approaches.

Another characteristic related to a dataset created by a single expert is the fact, that the entries are biased toward the individual experiences. As shown in Appendix B, the most frequent term in the newly added QA set is "fiber"[3]. It is coincidental that the data

[3] The most frequent terms in both datasets listed in Appendix B show a discrepancy between words overlapping in questions and answers. In the dialog-based scenario this is natural as the exchanged communication tends to be shorter and the answerer repeats the words from the questions less frequently.

creator had experienced many inquiries regarding this topic and it does not mean that this tendency will be mirrored in the future questions to the system. To tackle this problem, it is advisable to collect the users' inputs and continue expending our dataset after the system is deployed.

4.4 Using Out-of-Glossary Terms

To exemplify one more distinct difficulty of newly added QA pairs, in this subsection we describe some terms which do not exist in the glossary (and the regulatory itself) but were used by researchers in inquiries to the data creator during his career.

The regulatory documents as Export Trade Control Order, the Foreign Exchange Order and the Combined Matrix (which tabulates them) contain specific terms which often differ from the jargon used by researchers who are the target users of our expert system. For example, the term "re-entry object (or body)" which indicates a machine capable of returning to the ground from outer space is described as "re-entry vehicle" in the legal documents. Therefore, when the user uses "re-entry body" in the input, no related term will be matched. Similarly, "drone" become "unmanned aerial vehicle" and a "bearing" commonly written in Japanese in katakana (ベアリング - *bearingu*), are described in rarer Chinese ideograms (軸受 - *jikuuke*). Language models could be able to find similarity in above cases, but "soft matter" is an example which could be problematic. In usual context, this term is referring to flexible and tough gels in the chemical field, and new types of such materials are being created and announced on the regular basis. From the viewpoint of security export control, the focus is on the advanced materials specified in Paragraph 5 of the Export Trade Control Order, and also structural materials specified in Paragraph 4 and Paragraph 2 of the Order. For example, in the case of Paragraph 5, fibers and other materials specified in Article 4, Item 15 of the Ministerial Order Specifying Goods and Technologies, are to be considered. The specific modulus of elasticity and the specific strength of the material are used as a measure, but they are often lower than the standard values when viewed in terms of soft matter. The actual flow of the search-based reasoning algorithm would be as follows: soft matter → fiber → confirmation of specific elastic modulus and specific strength values. Such multi-hop approach is still difficult for question-answering systems and require step-by-step process with specific models using different algorithms.

Last example is "Raman scattering", also known as "Raman effect," which was discovered by the Indian physicist C.V. Raman and his student K.S. Krishnan in 1928, following Rayleigh scattering, known since the 19th century, and Mie scattering, discovered in 1908. In Rayleigh scattering, the scattered photon has the same energy (i.e., same frequency, wavelength, and color) as the incident photon. In contrast, Raman scattering has a different energy from the incident photon. Raman scattering itself as a physical phenomenon is not regulated, but it is used in applications such as material identification by using a laser as a monochromatic light source (Raman spectroscopy). In the context of security export control, Raman laser oscillators are regulated under Article 1, Item 36 of the Ministerial Order Specifying Goods and Technologies in the Paragraph 2 of Nuclear Power of the Export Trade Control Order. From the above explanation, it is not easy to label "Raman scattering" as *Applicable* or *Not applicable* without bringing it down to the level of "Raman laser oscillator". It is also impossible to directly find related passages in the regulatory texts and answering a question about this term could be difficult without referring to external knowledge sources.

As shown by the examples above, not only annotating such data is difficult; our new dataset can become a hard test for existing machine reading algorithms and reveal the shallowness of current machine learning approaches.

5. Conclusions and Future Work

In this paper we introduced an addition to our previous dataset for question answering for security export control expert system, which can be also used for experiments with standard methods for QA. Opposed to long questions and answers with many related keywords regarding mostly exceptions, the addition addresses short queries and answers made by the expert according to his work-experience by answering questions from academic researchers. This addition will allow to test our dialog-based expert system under development with more realistic input. The quantitative effect of adding typical QA to CISTEC's QA data will be confirmed in the next phase of our research. In the next step we are going to test several classifiers to test their capability to recognize types of questions and answer. Depending on the experimental results, we plan to implement the best model into our expert system for security export control. As approaches using language models without many examples like few-shot learning become more powerful also for Japanese language, we also want to test such methods in combination with multi-

hop reasoning required not only to find related passages from the regulatory documents, but also to generate explanations in natural language.

Acknowledgements

We would like to thank anonymous reviewers for their suggestions and comments. This work was supported by JSPS KAKENHI Grant Number 20K12556.

References

[1] Obayashi A. and Rzepka R., Annotated Question and Answer Dataset for Security Export Control. Proceedings of the 7th Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2021) IJCAI 2021 Workshop, CEUR Workshop Proceedings vol. 2935, Montreal, Canada, August 2021.

[2] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.

[3] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. 2019 Jul 26.

[4] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.

[5] Sovrano F, Palmirani M, Vitali F. Legal knowledge extraction for knowledge graph based question-answering. In Legal Knowledge and Information Systems 2020 (pp. 143-153). IOS Press.

[6] Abacha AB, Demner-Fushman D. Recognizing question entailment for medical question answering. In AMIA Annual Symposium Proceedings 2016 (Vol. 2016, p. 310). American Medical Informatics Association.

[7] Ashley KD. Automatically Extracting Meaning from Legal Texts: Opportunities and Challenges. Ga. St. UL Rev.. 2018;35:1117.

[8] Collarana D, Heuss T, Lehmann J, Lytra I, Maheshwari G, Nedelchev R, Schmidt T, Trivedi P. A question answering system on regulatory documents. In Legal Knowledge and Information Systems 2018 (pp. 41-50). IOS Press.

[9] Grabmair M, Ashley KD, Chen R, Sureshkumar P, Wang C, Nyberg E, Walker VR. Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools. In Proceedings of the 15th international conference on artificial intelligence and law 2015 Jun 8 (pp. 69-78).

[10] Tamper M, Oksanen A, Tuominen J, Hietanen A, Hyvönen E. Automatic annotation service appi: Named entity linking in legal domain. InEuropean Semantic Web Conference 2020 May 31 (pp. 208-213). Springer, Cham.

[11] Rzepka R, Shirafuji D, Obayashi A. Limits and Challenges of Embedding-based Question Answering in Export Control Expert System. Procedia Computer Science. 2021 Jan 1;192:2709-19.

Appendix A

Examples of additional questions and answers from the proposed additional QA set (translated to English by authors).

Question	Answer
ジルコニウムは輸出管理で規制されますか。	爆発物の反応材料で、ジルコニウム粉末や粉末からなる成型品が規制されます。 また、化学兵器の製造に用いられる装置で、反応器等でジルコニウムに裏打ちされたものも規制されます。その他、多くの用途で規制されます。
Is zirconium regulated by security export controls?	Reactive materials for explosives, such as zirconium powder or molded products made of the powder, are regulated. Also regulated are devices used in the manufacture of chemical weapons, such as reactors, lined with zirconium. Many other applications are also regulated.
ガス遠心分離機関連ではどういったものが規制されますか。	ガス遠心分離機のロータに用いられる構造材料や、ガス遠心分離機のロータに用いることができる構造材料等が規制されます。
What is regulated in relation to gas centrifuges?	Structural materials used in gas centrifuge rotors and structural materials that can be used in gas centrifuge rotors are regulated.
純チタンは輸出令別表第一に該当しますか。	純チタンであれば、少なくとも、輸出令別表第一の 1 項から 15 項までには該当しません。
Does pure titanium fall under Appended Table 1 of the Export Trade Control Order?	Pure titanium, at least, does not fall under paragraphs 1 through 15 of Appended Table 1 of the Export Trade Control Order.

輸出令別表第一の 3 項(1)に該当する貨物で、貨物等省令第二条第一項第一号から第三号の規定に「いずれかの物質の含有量が全重量の 30 パーセントを超えるもの」とありますが、これは含有量が全重量のちょうど 30 パーセントであれば該当しないと解釈してよいのですか。	そのように理解して結構です。
The provisions of Article 2, paragraph 1, items 1 through 3 of the Ministerial Order specifying Goods and Technologies, etc., for goods falling under paragraph 3 (1) of Appended Table 1 of the Export Trade Control Order state that "the content of any substance exceeds 30 percent of the total weight"; is this to be interpreted as meaning that if the content is exactly 30 percent of the total weight, it is not applicable?	It is fine to understand it that way.
潜水艇はどういったものが規制されますか。	1000 メートルを超える水深で使用できるように設計したものが規制されます。
What kind of submersibles are regulated?	Designed for use in water depths exceeding 1,000 meters are regulated.
人工衛星はどういったものが規制されますか。	軍用人工衛星に限らず、宇宙空間の飛翔体が広く規制されています。
What kind of satellites are regulated?	Not only military satellites, but also flying objects in space are widely regulated.
コンデンサはどういったものが規制されますか。	高電圧用のコンデンサで定格電圧が 5,000 ボルト以上のものが規制されます。また、パルスコンデンサで規制されるものがあります。
What capacitors are regulated?	Capacitors for high voltage applications with a rated voltage of 5,000 volts or more are regulated. There are also pulse capacitors that are regulated.
レーザー発振器とは何ですか。	輻射の誘導放出による光増幅を利用して空間的及び時間的にコヒーレントな光を発生させるものをいいます。（運用通達の解釈より）
What is a laser oscillator?	A device that generates spatially and temporally coherent light using optical amplification by induced emission of radiation. (From the interpretation of the Operational Notification)
レーザー発振器の部分品の注意事項は何ですか。	レーザー発振器の部分品については、他の用途に用いることができるものを含むので注意が必要です。
What are the precautions for laser oscillator components?	Note that partials of laser oscillators include those that can be used for other purposes.
レーザー発振器はどこで規制されますか。	広く規制されていますが、エキシマレーザー発振器、半導体レーザー発振器、化学レーザー発振器、一酸化炭素レーザー発振器、二酸化炭素レーザー発振器及び繰り返しでないパルスを発振するネオジムガラスレーザー発振器は、貨物等省令第９条第十号ニに基づいて判定してください。
Where are laser oscillators regulated?	Although widely regulated, excimer laser oscillators, semiconductor laser oscillators, chemical laser oscillators, carbon monoxide laser oscillators, carbon dioxide laser oscillators, and neodymium glass laser oscillators that emit non-repetitive pulses should be determined in accordance with Article 9, Item 10 (d) of the Ministerial Order Specifying Goods and Technologies, etc.
宇宙用に設計したとはどういう意味ですか。	地表から１００キロメートルを超える高度で動作する設計若しくは製造されたこと又は試験に合格して適格とされたことをいいます。（運用通達の解釈より）
What do you mean by "designed for space"?	Designed or manufactured to operate at altitudes exceeding 100 kilometers above the earth's surface, or qualified by passing a test. (From the Interpretation of the Operational Notification)
アルミは規制されますか。	アルミニウム合金や粉末が単体で規制されている他、アルミニウム化合物である窒化アルミニウム等、幅広く規制されています。
Is aluminum regulated?	Aluminum alloys and powders are regulated as stand-alone products, and aluminum compounds, such as aluminum nitride, are also widely regulated.

Appendix B

1) Most common (left table) and least common (right) 15 terms in the CISTEC dataset **questions**.

判定	decision
使用	usage
機能	function
測定	measurement
プログラム	program
機械	machine
制御	control
部分	part
技術	technology
対象	object
許可	permission
承認	approval
規定	regulation
電子	electronic
回路	circuit

画像	image
ふっ素	fluorine
プラント	plant
プロセス	process
製剤	product
放射	radiation
単一	single
実現	realization
イオン	ion
公開	public
デバイス	device
コア	core
加重	weighting
プロセッサ	process
意図	intention

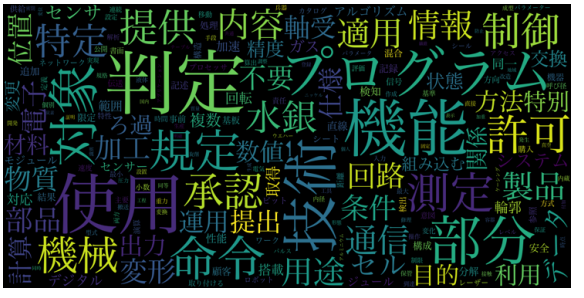
2) Most common (left table) and least common (right) 15 terms in the CISTEC dataset **answers**.

判定	decision
使用	usage
機能	function
技術	technology
プログラム	program
部分	part
対象	object
命令	order
規定	regulation
測定	measurement
許可	permission
機械	machine
承認	approval
制御	control
提供	offer

変位	displacement
表示	indication
識別	discrimination
付加	addition
地域	area
国内	national
アルミニウム	aluminum
共通	common
工程	production process
個人	individual
質量	mass
加重	weighting
明確	clarity
炭素	carbon
繊維	fiber

3) Word cloud of most common **questions** (CISTEC)

4) Word cloud of most common **answers** (CISTEC)



5) Most common (left) and least common (right) 15 terms in the proposed additional dataset **questions**.

繊維	fiber
炭素	carbon
許可	permission
部品	part
ガス	gas
窒素	nitrogen
ジルコニウム	zirconium
原子	atom
外径	outside diameter
混合	mixture
判定	decision
使用	usage
ころ	(parser noise)
ロボット	robot
レーザー	laser

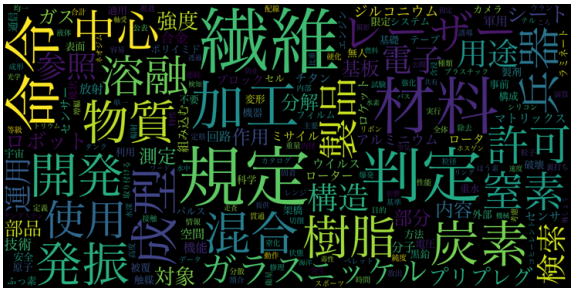
投下	drop
レジスト	regist
設置	installation
加工	manufacturing
パラメータ	parameter
流体	liquid
病気	illness
海洋	ocean
プレート	plate
レート	rate
相互	mutual
作用	operation
処理	processing
スラブ	slab
改善	improvement

6) Most common (left) and least common (right) 15 terms in the proposed additional dataset **answers**.

繊維	fiber
規定	regulation
材料	material
判定	decision
命令	order
加工	manufacturing
成型	casting
炭素	carbon
樹脂	resin
混合	mixture
開発	development
物質	substance
発振	oscillate
兵器	weapon
溶融	meltdown

光学	optics
走査	scan
透過	permeation
公開	public
ホスゲン	phosgene
毒性	toxicity
リング	ring
軸受	bearing
ワード	word
飛しょう	flight
バス	bus
比重	specific gravity
空気	air
検知	detection
水中	underwater

7) Word cloud of most common **questions** (proposed)



8) Word cloud of most common **answers** (proposed)