

PhotoshopQuiA: A Corpus of Non-Factoid Questions and Answers for Why-Question Answering

Andrei Dulceanu^{†*}, Thang Le Dinh[‡], Walter Chang^{*}, Trung Bui^{*}, Doo Soon Kim^{*},
Manh Chien Vu[‡], Seokhwan Kim^{*}

[†]Universitatea Politehnica București, România

^{*} Adobe Research, California, United States

[‡]Université du Québec à Trois-Rivières, Québec, Canada

andrei.dulceanu@cs.pub.ro, {wachang, bui, dkim, seokim}@adobe.com, {thang.ledinh, manh.chien.vu}@uqtr.ca

Abstract

Recent years have witnessed a high interest in non-factoid question answering using Community Question Answering (CQA) web sites. Despite ongoing research using state-of-the-art methods, there is a scarcity of available datasets for this task. Why-questions, which play an important role in open-domain and domain-specific applications, are difficult to answer automatically since the answers need to be constructed based on different information extracted from multiple knowledge sources. We introduce the PhotoshopQuiA dataset, a new publicly available set of 2,854 why-question and answer(s) (WhyQ, A) pairs related to Adobe Photoshop usage collected from five CQA web sites. We chose Adobe Photoshop because it is a popular and well-known product, with a lively, knowledgeable and sizable community. To the best of our knowledge, this is the first English dataset for Why-QA that focuses on a product, as opposed to previous open-domain datasets. The corpus is stored in JSON format and contains detailed data about questions and questioners as well as answers and answerers. The dataset can be used to build Why-QA systems, to evaluate current approaches for answering why-questions, and to develop new models for future QA systems research.

Keywords: question answering, community question answering, non-factoid question, Why-QA

1. Introduction

The success of IBM’s Watson in the *Jeopardy!* TV game-show in 2011 and the significant investments of large tech companies in building personal assistants (e.g., Microsoft Cortana, Apple Siri, Amazon Alexa or Google Assistant) have strengthened the interest in the Question Answering field. These systems have in common the fact that they mostly tackle **factoid questions**. These are questions that “can be answered with simple facts expressed in short text answers”; usually, their answers include “short strings expressing a personal name, temporal expression, or location” (Jurafsky and Martin, 2017). An example of a factoid question and its answer is:

Q: *Who is Canada’s prime minister?*

A: *Justin Trudeau.*

By contrast, **non-factoid questions** ask for “opinions, suggestions, interpretations and the like” (Tomasoni and Huang, 2010). Answering and evaluating the quality of the provided answers for non-factoid questions have proved to be non-trivial due to the difficulty of the task complexity as well as the lack of training data. To address the latter issue, numerous researchers have tried to take advantage of user-generated content on Community Question Answering (CQA) web sites such as Yahoo! Answers¹, Stack Overflow² or Quora³. These web forums allow users to post their own questions, answer others’ questions, comment on others’ replies, and upvote or downvote answers.

Usually, if a user is the original questioner, he/she is allowed to select the most relevant answer to his/her question. Although CQA web sites have lots of experts, it still takes their time to give pertinent, authoritative answers to user questions and not all the content shares the same characteristics. Some key differences (Blooma and Kurian, 2011) in answer quality and availability between traditional QA systems and CQA web sites include: the type of questions (factoid vs. non-factoid), the quality of the answers (high vs. varying from answerer to answerer) and the response time (immediate vs. several hours or days).

Among all categories of non-factoid questions, namely list, confirmation, causal and hypothetical (Mishra and Jain, 2016), we are especially interested in why-questions that are related to causal relations. Why-questions are difficult to answer automatically since the answers often need to be constructed based on different information extracted from multiple knowledge sources. For this reason, why-questions need a different approach than factoid questions because their answers usually cannot be stated in a single phrase (Verberne et al., 2010).

A why Question Answering (Why-QA) system trying to answer questions using CQA data needs to be able to distinguish between relevant and irrelevant answers (answer selection task). Most of the time these systems also produce a sorted output of relevant answers (answer re-ranking task). Both tasks require curated and informative datasets on which to evaluate proposed methods.

In this paper, we introduce the PhotoshopQuiA dataset, a corpus consisting of 2,854 (WhyQ, A) pairs covering various questions and answers about Adobe Photoshop⁴. We

¹<https://answers.yahoo.com>

²<https://stackoverflow.com>

³<https://www.quora.com>

⁴Adobe Photoshop is the *de facto* industry stan-

chose Adobe Photoshop because it is a popular and well-known product, with a lively, knowledgeable and sizable QA community. PhotoshopQuiA is the first large Why-QA only English dataset that focuses on a product, as opposed to previous open-domain datasets. Our dataset focuses on why-questions that occur while a user is trying to complete a task (e.g., changing color mode for an image, or applying a filter). It contains contextual information about the answer, which in turn makes it easier to build a QA system that is able to find the most relevant answers. We named the corpus PhotoshopQuiA, because *quia* (first and last letters capitalized as in Question Answering) means *because* in Latin and therefore hints at the expected why-answer type.

One of the challenges that often arises with CQA data is the high variance in quality for both questions and answers. To address this problem, we include in our dataset mostly official answers (65.5% from total pairs) given by Adobe Photoshop experts. We choose to provide both text and HTML representations of questions and answers because the raw HTML snippets often include relevant information like documentation links, screenshots or short videos which would be otherwise lost. We analyze the (WhyQ, A) pairs for presence of certain linguistic cues such as causality markers (e.g., *the reason for*, *because* or *due to*).

2. Related Work & Datasets

In recent years, numerous datasets have been released in the domain of question-answering (QA) systems to promote new methods that integrate natural language processing, information retrieval, artificial intelligence and knowledge discovery. The majority of these datasets were open-domain (Bollacker et al., 2008; Ignatova et al., 2009; Cai and Yates, 2013; Yang et al., 2015; Chen et al., 2017). There are still a few QA datasets for specific fields such as BioASQ and WikiMovies. The BioASQ dataset contains questions in English, along with reference answers constructed by a team of biomedical experts (Tsatsaronis et al., 2015). The WikiMovies dataset contains 96K question-answer pairs in the domain of movies (Miller et al., 2016). Table 1 introduces selected recent open-domain QA datasets.

Several existing datasets focus on the data taken from CQA web sites. The data structure of our dataset (question-answer(s) pair with the best answer labeled), resembles the one in (Hoogeveen et al., 2015). However, it does not include comments and tags, making it more suitable for Why-QA than previous structures which include only questions (Iyer et al., 2017). Our work is mostly related to the SemEval-2016 Task 3 dataset (Nakov et al., 2016) which contains more than 2,000 questions associated with ten most relevant comments (answers). It also shares some characteristics with Yahoo’s Webscope⁵ L4 used by (Surdeanu et al., 2008) and (Jansen et al., 2014), L6 and with

standard raster graphics editor developed and published by Adobe Systems for macOS and Windows.
https://en.wikipedia.org/wiki/Adobe_Photoshop
⁵<https://webscope.sandbox.yahoo.com/catalog.php?datatype=1>

nfL6⁶, a non-factoid subset of L6 focusing only on how-questions (Cohen and Croft, 2016). Although Yahoo Webscope L6 certainly contains many why-QA pairs which should be fairly trivial to extract from the entire dataset, we believe this limits its usefulness for building Why-QA systems. While all these datasets focus on CQA forums data, there are some key differences between our work and existing datasets (Table 2).

Dataset	Description
WebQuestions and Free917	for training semantic parsers, which map natural language utterances to denotations (answers) via intermediate logical forms (Berant et al., 2013)
CuratedTREC	2,180 questions extracted from the datasets from TREC (Baudiš and Šedivý, 2015)
WikiQA	3,000 questions sampled from Bing query logs associated with a Wikipedia page presumed to be the topic of the question (Yang et al., 2015)
30M Factoid QA Corpus	30M natural language questions in English and their corresponding facts in the knowledge base Freebase (Serban et al., 2016)
SQuAD	100,000 question-answer pairs on more than 500 Wikipedia articles (Rajpurkar et al., 2016)
Amazon	1.4 million answered questions from Amazon (Wan and McAuley, 2016)
Baidu	42K questions and 579K evidences, which are a piece of text containing information for answering the question (Li et al., 2016)
Allen AI Science Challenge	2,500 questions. Each question has 4 answer candidates (Chen et al., 2017)
Quora	Over 400K sentence pairs of which, almost 150K are semantically similar questions; no answers are provided (Iyer et al., 2017)

Table 1: Recent datasets for QA systems

Difference	This work	Previous work
Scope	closed domain (focus on product usage)	open domain
Answer authority	picked by a domain expert (65.5%)	n/a
Question types	why-questions only	various (focus on how-questions)

Table 2: Major differences between our dataset and previous datasets focusing on CQA forums data

The difference in scope allows researchers to verify

⁶<https://ciir.cs.umass.edu/downloads/nfL6>

whether all previous research achievements made on open domain datasets such as Yahoo Webscope could be translated into a closed domain such as ours, or whether domain adaptation is needed. The authors are aware of the corpora from (Prasad et al., 2007) and (Dunietz et al., 2017), but these were not considered for this work, because their datasets do not address CQA and/or Why-QA.

Some of the previous studies in Why-QA systems tried to extract why-questions from QA datasets related to general questions; however, the size and quality of why-questions were limited. Previous datasets used in Why-QA task contain few (WhyQ, A) pairs (under 1,000), are handcrafted, are not available online anymore (Verberne et al., 2007; Mrozinski et al., 2008; Higashinaka and Isozaki, 2008) or target Japanese (Higashinaka and Isozaki, 2008; Oh et al., 2012). There is a need for a public specific why-question dataset for English to advance the research and development in Why-QA.

3. PhotoshopQuiA Dataset

In this section we describe the process of creating the PhotoshopQuiA dataset and succinctly compare our data collection approach to previous related approaches.

3.1. Data Sources

We identified the following five web sites as appropriate sources for collecting why-questions about Adobe Photoshop: Adobe Forums ⁷, Stack Overflow, Graphic Design ⁸, Super User ⁹ and Feedback Photoshop ¹⁰. Although there were additional CQA web sites containing Photoshop-related questions and answers, we selected only the sources above because they all have moderated, recent, high-quality and authoritative content. Regarding the last two points, it is worth mentioning that the dataset contains a high ratio of answers coming from Adobe experts working in the Photoshop team (65.5% of total answers).

When using one of the above-mentioned forums, the original questioner has the possibility to select the most relevant answer to his/her question. This is often referred as the accepted answer. If such an answer does not exist, does not fully meet established criteria or even does not solve the problem at hand, registered users may upvote or downvote it. As stated previously, PhotoshopQuiA includes all answers available for each why-question, labeling the correct answer. We strove to always label as correct accepted answers only, but when such answers were not available, we selected the most voted answer instead. If the most voted answer had at least one downvote, we did not include the (WhyQ, A) pair altogether.

Our approach resembles previous datasets described in related work section, where a dataset item contained either the full conversation thread, (Hoogeveen et al., 2015)(although we did not include tags or comments), or multiple relevant answers for a question (Nakov et al., 2016). One of the key advantages of our approach is that our answers

are more reliable and correct - two thirds of the (WhyQ,A) pairs have an accepted answer authored by a Photoshop expert.

3.2. Web Crawling

We used Scrapy ¹¹, an application framework written in Python for our web scraping task. Web scraping includes two main stages: web crawling (i.e., fetching or downloading a web page) and data extraction (i.e., extracting structured content from a fetched page). In order to successfully scrape a CQA web site, Scrapy needs a `Spider` definition containing the following:

- an initial list of URLs which the `Spider` will begin to crawl from. This is provided in `start_urls` attribute or via `start_requests()` method.
- an implementation of the default `parse()` callback method, which is a generator function under the hood. This method is accountable for all the heavy lifting needed for processing the response corresponding to each request. When all content is available on the requested web page, it yields a Python dictionary filled with data of interest. If additional requests need to be performed (i.e. following a new URL found in the page), it yields a new request which in turn needs
- an implementation of a new, custom, user defined callback method for parsing the new response. This method will yield the final scraped items.

All Scrapy requests are processed and scheduled asynchronously, enabling fast concurrent crawls. In order to politely scrape mentioned CQA web sites, we overrode Scrapy default settings and limited the number of concurrent requests per IP to one, with a five seconds download delay between each request.

After taking a look at the building blocks of a Scrapy `Spider`, few words should be mentioned about how the actual item scraping works in the `parse()` callback. This is done in two phases: selection and extraction. Selection employs CSS selectors for selecting HTML elements in the response. Once the desired elements are selected, XPath expressions can be used for a more fine-grained control of the extracted content.

3.3. Data Collection

The same steps described below were followed for each CQA web site considered. For brevity, we only describe the full workflow used for scraping Stack Overflow. We first manually performed a search containing “why Photoshop” keywords, quotes excluded. In the next step, we used resulting URL and number of result pages to handcraft the `start_urls` list. For example, performing a search on Stack Overflow using our search query and clicking the first result page at the bottom resulted in the following link: <https://stackoverflow.com/search?page=1&tab=Relevance&q=why%20Photoshop>. We iterated over the number of result pages returned by the search

⁷<https://forums.adobe.com/welcome>

⁸<https://graphicdesign.stackexchange.com>

⁹<https://superuser.com>

¹⁰<https://feedback.photoshop.com>

¹¹<https://scrapy.org>

and included the appropriate URL (page number changed) in `start_urls`.

In the `parse()` method we iterated over each search result on the page and created a new `WhyQuestionAnswerPair` item. This item along with the actual URL of the search result were passed in turn to `parse_question_answer_pair()` callback. This method extracted all the information needed from the new URL and returned the populated item. Finally, each result item was appended to the consolidated JSON file.

3.4. Annotation

We retain the following data and meta-data for each (WhyQ, A) pair: question identifier on the source web site, question URL, question title, question date, questioner, questioner level, question state (open or resolved), full question text, full question HTML and for each answer available: answer date, answerer, answerer level, answer votes, answer text, full answer HTML and a boolean property indicating if this is the best answer or not. Beautiful Soup¹² was used for extracting question and answer(s) text from HTML; links provided inside raw snippets were enclosed in parentheses and kept in the final text excerpt.

Since almost all previous properties are self-explanatory we will give more insight into questioner/answerer level properties. Depending on the source web site these properties can either be a text describing user seniority/level on the web site (e.g., “Level 1”, “Adobe Community Professional”, etc. for Adobe Forums), a number describing the number of posts written so far (Feedback Photoshop) or the reputation score of the user (Stack Overflow and the like). We treat question identifier, questioner level, answerer level and answer votes as optional properties. All other properties not listed as optional are required. We enforce these constraints and validate our dataset by using a JSON schema¹³. A sample (WhyQ,A) pair from our corpus is available in the appendix.

3.5. Post-Processing

After the web scraping phase, we ended up with a consistently larger number (4,365 pairs obtained) of (WhyQ,A) answer pairs than those included in our final dataset (2,854 pairs). We went further and refined these pairs in a two-stage manual cleanup process in which we removed duplicates, questions which mentioned Adobe Photoshop incidentally and questions containing different question types (e.g., “how to”, “definition”) than our intended why-question type.

3.6. Availability & Reuse

The dataset, its JSON schema and a copy of this paper are available on *GitHub*¹⁴ under cc by-sa 3.0 with attribution required license¹⁵. We also plan to add to that repository

¹²<https://www.crummy.com/software/BeautifulSoup>

¹³<http://json-schema.org>

¹⁴<https://github.com/dulceanu/photoshop-quia>

¹⁵<https://creativecommons.org/licenses/by-sa/3.0/>

the Python scripts used for web scraping and manipulating the data.

4. Data Statistics & Analysis

The statistics of the PhotoshopQuiA dataset are described in Table 3. When crunching the data we ignored 13 (WhyQ,A) pairs with long question content. The questioners posted OS- and Photoshop-related configurations which artificially increased the numbers.

Metric	Value
Avg. no. of words in question title	10.69
Avg. no. of words in question text	102.29
Avg. no. of words in best answer	95.60

Table 3: Statistics of the PhotoshopQuiA dataset

Causality, causation or **causal relation** is “the relation between a cause and its effect or between regularly correlated events or phenomena”¹⁶. Words connecting the cause and effect parts of a causal relation are called **causal markers**. As outlined by (Girju, 2003), causative constructions can be explicit (introduced by causal markers like *cause*, *effect*, *consequence*, etc.) or implicit (i.e., without any explicit marker). Blanco et al. (2008) refine causal relations categories by introducing ambiguity (if the causal marker does not always signal a causation, e.g., *since*) and completeness (when both cause and effect parts are present).

For our analysis, we are interested in explicit causal markers, ignoring ambiguity and completeness of the causal relations. To come up with an extended list of causal markers, we started with the adverb *therefore* and looked for its synonyms in the Oxford Thesaurus of English. We found that causality markers are present in 1,583 answers out of 2,854 total answers (55.46%). Figure 1 shows the distribution of the 22 markers found. A future study needs to be done to remove the ambiguous markers.

To better visualize the diversity of the questions in the dataset, we include in Figure 2 a breakdown of questions based on the first wh-word contained in question title.

When crunching the numbers behind these statistics, we noticed that 239 questions (8.37% of the dataset) have multiple wh-words in their title. A distinct category emerged from questions asking for OS specific details or solutions, totaling 286 questions, or 10.26% of the dataset. Moreover, 763 questions (accounting for 26.73% of the total questions) contained the adverb *not*. The key takeaway here is that the two most important features when it comes to question titles are the presence of the word *why* and of the negation. At first, it might seem odd that 9.76% of the questions with at least one wh-word contain the word *how* and are still considered why-questions. An example of such a question is:

Q Title: *Photoshop: how to change from RGB to CMYK without any color loss?*

Q Content: *[...] When converting this from RGB to*

¹⁶From Merriam-Webster’s Online Dictionary

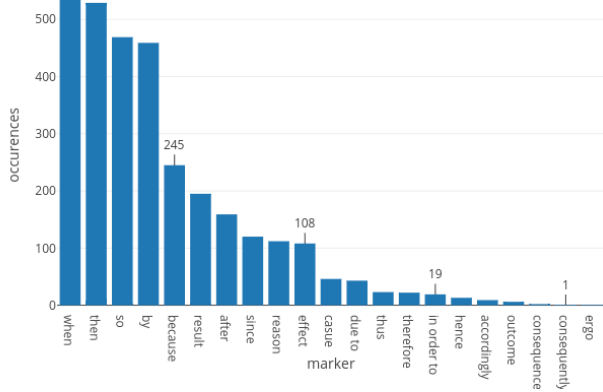


Figure 1: Distribution of explicit causality markers found in PhotoshopQuiA (1,583 answers)

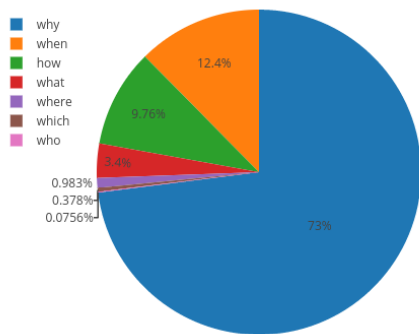


Figure 2: Distribution of questions based on first occurrence of a wh-word in question title (1,322 questions)

CMYK, my original colors are changing. [...]
A: By changing color mode you essentially change colors. [...]
 [...] Because of that, [...] the colors you get [...] will be made as close to original as possible - but not identical.[...]

As it can be seen, the questioner was not interested in a general recipe about changing the color mode of his/her document, but rather wanted to do it under special circumstances (i.e. without color loss). In other words, this question is equivalent to a why-question like *Why do I have a color loss when changing from RGB to CMYK?*. Often, people use *how*, usually followed by *come*, to informally ask for causes of events, as outlined by this example from Merriam-Webster's Online Dictionary: *How come you can't go?*.

A detailed classification of the questions containing the word *why* in their title is included in Figure 3. The num-

bers are presented again in percentages and the categories are not mutually exclusive. In orange we present the percent of why-questions in that sub-category which have also the adverb *not* in their title. Noteworthy here is the strong connection between the modal *can* and the negation (e.g., *Why can't I amend my mask with the brush tool?*) and between the auxiliary *will* and the negation (e.g., *Why won't Photoshop let me rename my layers?*).

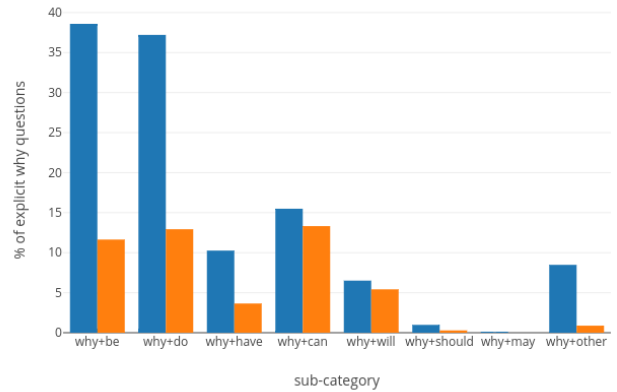


Figure 3: Explicit why questions sub-categories (1,006 questions)

5. Expected Use

Below we suggest some use cases of tasks that we are currently working on:

- Question Classification: what classification models for why-questions can be used to facilitate the tasks of Why-QA systems and how to classify the why-questions based on the selected classification model.
- Semantic parsing: how to convert a why-question into a structured format (or a canonical form) to speed up the matching process.
- Question to question matching: what approaches can be used to match the asked question and the questions in the knowledge base (information retrieval, bag-of-word, knowledge based approach, or neural network).

6. Conclusion and Future Work

In this paper, we present PhotoshopQuiA, a new dataset for Why-QA. The dataset is constructed in a natural and practicable manner so that it can be used to observe different characteristics and behaviors of the why-questions.

We believe that the PhotoshopQuiA dataset enables new research in Why-QA, which has received less focus, and that our work stimulates further research to advance the QA technology needed for smart services such as recommendation systems, chatbots, and intelligent assistants.

7. Acknowledgements

The authors express their sincere thank to the University Gift Funding of Adobe Systems Incorporated for the partial financial support for this research.

Appendix: JSON format of the PhotoshopQuIA dataset

```
1 "id": 20518335,
2 "url": "https://stackoverflow.com/questions/20518335/why-do-photoshop-files-start
  -with-8bps",
3 "title": "Why do Photoshop files start with 8BPS?",
4 "questionDate": "2013-12-11T11:48:22Z",
5 "questioner": "Party Ark <https://stackoverflow.com/users/548664/party-ark>",
6 "questionerLevel": "532",
7 "questionState": "resolved",
8 "questionText": "From pretty much the beginning of time Photoshop files have
  begun with 8BPS. (I have verified this back to version 2.5) It must have had
  some meaning at some point.\nThe 8B I thought might refer to bits/channel, but
  it makes no difference saving it 16 or 32. PS is probably PhotoShop, but
  might not be. Something to do with the way Mac saved files?",
9 "questionHtml": "<div class=\"post-text\" itemprop=\"text\">\r\n\r\n<p>From
  pretty much the beginning of time Photoshop files have begun with 8BPS. (I
  have verified this back to version 2.5) It must have had some meaning at some
  point.</p>\n\n<p>The 8B I thought might refer to bits/channel, but it makes no
  difference saving it 16 or 32. PS is probably PhotoShop, but might not be.
  Something to do with the way Mac saved files? </p>\n  </div>",
10 "answers": [
11   {
12     "answerDate": "2015-02-20T22:31:45Z",
13     "answerer": "Community <https://stackoverflow.com/users/-1/community>",
14     "answererLevel": "1",
15     "answerVotes": 4,
16     "answerText": "8B is shorthand for Adobe. I guess \"eight bee\" sounds a
      little bit like Adobe; more so if you're Italian - \"Otto Bee\".\nAnd PS
      is \"Photoshop\". So, 8BPS is \"Adobe Photoshop\".\n8B crops up in
      quite a few places in Adobe file extensions or internal types. Wikipedia
      has a list (http://en.wikipedia.org/wiki/8B).",
17     "answerHtml": "<div class=\"post-text\" itemprop=\"text\">\r\n<p><code>8B</
      code> is shorthand for Adobe. I guess \"eight bee\" sounds a <em>little
      </em> bit like Adobe; more so if you're Italian - \"Otto Bee\".</p>\n\n<
      p>And <code>PS</code> is \"Photoshop\". So, <code>8BPS</code> is \"Adobe
      Photoshop\".</p>\n\n<p>8B crops up in quite a few places in Adobe file
      extensions or internal types. <a href=\"http://en.wikipedia.org/wiki/8B
      \" rel=\"nofollow\">Wikipedia has a list</a>.</p>\n  </div>",
18     "bestAnswer": true
19   },
20   {
21     "answerDate": "2013-12-11T11:53:19Z",
22     "answerer": "Syjin <https://stackoverflow.com/users/733368/syjin>",
23     "answererLevel": "1,922",
24     "answerVotes": 0,
25     "answerText": "That is just the Signature to identify the file as a
      Photoshop file. From the Specification:\nSignature: always equal to '8
      BPS' . Do not try to read the file if the signature does not match this
      value.\nSee the Photoshop File Format Specification (http://www.adobe.
      com/devnet-apps/photoshop/fileformatashtml/#50577409\_pgfid-1036097) for
      more detailed information.",
26     "answerHtml": "... truncated due to space constraints...",
27     "bestAnswer": false
28   }
29 ]
30 }
```

Listing 1: An example of an item containing all properties from PhotoshopQuIA dataset

8. Bibliographical References

- Baudiš, P. and Šedivý, J. (2015). Modeling of the question answering task in the yodaqa system. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 222–228. Springer.
- Berant, J., Chou, A., Frostig, R., and Liang, P. (2013). Semantic parsing on freebase from question-answer pairs. In *EMNLP*, volume 2, page 6.
- Blanco, E., Castell, N., and Moldovan, D. I. (2008). Causal relation extraction. In *Lrec*.
- Blooma, M. J. and Kurian, J. C. (2011). Research issues in community based question answering. In *PACIS*, page 29.
- Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. AcM.
- Cai, Q. and Yates, A. (2013). Large-scale semantic parsing via schema matching and lexicon extension. In *ACL (1)*, pages 423–433.
- Chen, D., Fisch, A., Weston, J., and Bordes, A. (2017). Reading wikipedia to answer open-domain questions. *arXiv preprint arXiv:1704.00051*.
- Cohen, D. and Croft, W. B. (2016). End to end long short term memory networks for non-factoid question answering. In *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval*, pages 143–146. ACM.
- Dunietz, J., Levin, L., and Carbonell, J. (2017). The because corpus 2.0: Annotating causality and overlapping relations. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 95–104.
- Girju, R. (2003). Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*, pages 76–83. Association for Computational Linguistics.
- Higashinaka, R. and Isozaki, H. (2008). Corpus-based question answering for why-questions. In *IJCNLP*, pages 418–425.
- Hoogeveen, D., Verspoor, K. M., and Baldwin, T. (2015). Cqadupstack: A benchmark data set for community question-answering research. In *Proceedings of the 20th Australasian Document Computing Symposium*, page 3. ACM.
- Ignatova, K., Toprak, C., Bernhard, D., and Gurevych, I. (2009). Annotating question types in social q&a sites. In *Tagungsband des GSCL Symposiums 'Sprachtechnologie und eHumanities*, pages 44–49.
- Iyer, S., Dandekar, N., and Csernai, K. (2017). First quora dataset release: Question pairs.
- Jansen, P., Surdeanu, M., and Clark, P. (2014). Discourse complements lexical semantics for non-factoid answer reranking. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 977–986.
- Jurafsky, D. and Martin, J. H. (2017). *Speech and language processing*. Unpublished Draft, 3rd edition.
- Li, P., Li, W., He, Z., Wang, X., Cao, Y., Zhou, J., and Xu, W. (2016). Dataset and neural recurrent sequence labeling model for open-domain factoid question answering. *arXiv preprint arXiv:1607.06275*.
- Miller, A., Fisch, A., Dodge, J., Karimi, A.-H., Bordes, A., and Weston, J. (2016). Key-value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126*.
- Mishra, A. and Jain, S. K. (2016). A survey on question answering systems with classification. *Journal of King Saud University-Computer and Information Sciences*, 28(3):345–361.
- Mrozinski, J., Whittaker, E., and Furui, S. (2008). Collecting a why-question corpus for development and evaluation of an automatic qa-system. *Proceedings of ACL-08: HLT*, pages 443–451.
- Nakov, P., Márquez, L., Moschitti, A., Magdy, W., Mubarak, H., Freihat, a. A., Glass, J., and Randeree, B. (2016). Semeval-2016 task 3: Community question answering. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 525–545, San Diego, California, June. Association for Computational Linguistics.
- Oh, J.-H., Torisawa, K., Hashimoto, C., Kawada, T., De Saeger, S., Kazama, J., and Wang, Y. (2012). Why question answering using sentiment analysis and word classes. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 368–378. Association for Computational Linguistics.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., and Webber, B. L. (2007). The penn discourse treebank 2.0 annotation manual.
- Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.
- Serban, I. V., García-Durán, A., Gulcehre, C., Ahn, S., Chandar, S., Courville, A., and Bengio, Y. (2016). Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus. *arXiv preprint arXiv:1603.06807*.
- Surdeanu, M., Ciaramita, M., and Zaragoza, H. (2008). Learning to rank answers on large online qa collections. *Proceedings of ACL-08: HLT*, pages 719–727.
- Tomasoni, M. and Huang, M. (2010). Metadata-aware measures for answer summarization in community question answering. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 760–769. Association for Computational Linguistics.
- Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M. R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., et al. (2015). An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. *BMC bioinformatics*, 16(1):138.
- Verberne, S., Boves, L., Oostdijk, N., and Coppen, P.-A. (2007). Evaluating discourse-based answer extraction for why-question answering. In *Proceedings of the*

- 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 735–736. ACM.
- Verberne, S., Boves, L., Oostdijk, N., and Coppen, P.-A. (2010). What is not in the bag of words for why-qa? *Computational Linguistics*, 36(2):229–245.
- Wan, M. and McAuley, J. (2016). Modeling ambiguity, subjectivity, and diverging viewpoints in opinion question answering systems. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 489–498. IEEE.
- Yang, Y., Yih, W.-t., and Meek, C. (2015). Wikiqa: A challenge dataset for open-domain question answering. In *EMNLP*, pages 2013–2018.