Recommendation of documents in the legal domain

Leonardo Rossi*, Renan Yamaguti†- FEEC‡

December 2021

Abstract

This work is our Final Project for the class EA376E. Since 1966, all public Brazilian acts of the Executive, Legislative and Judiciary Powers must be published in official journals to be legally valid [3]. However, there is no current standard for the publication of these journals. In order to provide a data center with official journals published in several cities for Brazilian society in general, Open Knowledge Brasil (OKBR) [1] created the project "Querido Diário" [7] to map, collect, process and make them available online. One of the current goals of the "Querido Diário" project is to provide content recommendations based on the searches and usage of each user. This work proposes a recommendation system for the "Querido Diário" project, using BERTimbau's [11] embeddings as inputs to a KNN [5, 6] classifier.

1 Introduction

In Brazilian law, under Law 4,965, of 1966, May 5th, any public act relating to employees of centralized administration bodies and autarchies, whether from the Executive, Legislative or Judiciary, will only have legal validity upon publication in an official journal [3]. However, there is no standardization of official journals in Brazil or a data center that allows civil society to access the content of journals easily and quickly. Given the lack of standardization of municipal daily data, Open Knowledge Brasil (OKBR) [1] created the project "Querido Diário", whose objective is to map, collect and process all municipal official journals in the country and make them available in an open format for civil society [7]. To carry out these three tasks, Open Knowledge Brasil has a team of scientists and data engineers and a large community of volunteer collaborators.

Currently, almost all official journals of 5,570 Brazilian cities have already been mapped by the "Querido Diário" community, with information on where

*l261900@dac.unicamp.br

†r262731@dac.unicamp.br

‡Faculty of Electrical and Computer Engineering (FEEC) - Unicamp - Campinas / Brazil

these journals are published, their temporal coverage and the type of file in which these journals are published. There are also journal scrapers for more than 2,300 cities, around 41% of the total. This data is made publicly available via the open API available at [9]. Due to the need to verify the data collected, only journals from 13 Brazilian capital cities are available in the API.

One of "Querido Diário" goals is to make content recommendations to a user when he searches in "Querido Diário". Ideally, the most relevant result should be shown first, on the main screen, but other official journals can also be recommended from temporal cuts (e.g. what are other similar journals published on the same day) or geographic cuts (e.g. what are other cities have similar content).

Thus, our work[1] proposes a recommendation system for official journals content available in "Querido Diário". For tokenization of the input data, we use the tokenizer of the trained agent in BERTimbau [11] (BERT [4] agent trained in Brazilian Portuguese). We use BERTimbau's embeddings to feed a KNN classifier, which is used in the semantic content recommendation task.

2 Methodology

We propose a recommendation system for "Querido Diário". For this, we collect PDFs available from "Querido Diário" directly through online links. With the collected data, we use pre-processing techniques to remove unwanted words, punctuation, numeration and to divide each PDF into paragraphs. Each paragraph is converted to tokens using BERTimbau's tokenizer. We got the embeddings produced by BERTimbau from these tokens. With these embeddings, we feed a KNN classifier to carry out a recommendation for one of the documents based on another. Our flowchart process is shown in Figure 2.

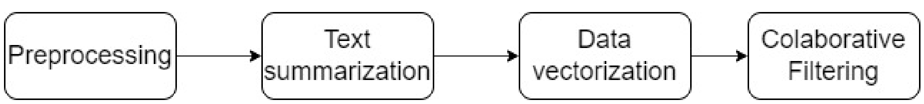


Figure 1: Our project's flowchart.

2.1 Preprocessing

The PDFs available in "Querido Diário" have text mixed with images and tables. Some official journals were only uploaded with photos from the official journal. Thus, it was necessary to establish some pre-processing functions, removing images and tables, removing small numbers, excluding accents and dividing the text into paragraphs. A list of stop-words in Brazilian Portuguese was defined to remove words that are commonly used in any text (prepositions, articles, etc).

[1] Our code is available at https://github.com/leolellisr/npl_recom_querido_diario

2.2 BERT and BERTimbau

The BERT model [4], which uses the Transformer architecture, despite having a multilingual model trained in 104 languages, also has monolingual models and BERT derivatives in single languages. The Brazilian Portuguese version of BERT, called BERTimbau [11], was trained in two sizes: base (with 12 layers, 768 hidden dimensions, 12 attention heads and 110M parameters) and large (with 24 layers, 1024 hidden dimensions, 16 attention heads and 330M parameters). The maximum length of a phrase is 512 tokens.

In our project, the following models were used: BERT ('bert-base-uncased') and BERTimbau ('neuralmind/bert-base-portuguese-cased'). Due to the architectures of each of the models, the inputs used to collect the embeddings had to be divided into every 512 tokens and then concatenated.

2.2.1 Tokenization

In token-level tasks, the document's context is used for input examples rather than sentence context to take advantage of longer contexts when encoding BERT token representations [4]. Examples larger than S tokens are divided into extents of length up to S using a pass of D tokens. Each extension is used as a separate example during training.

In BERT and BERTimbau tokenization, the original word is split into smaller subwords and characters. This is because BERT and BERTimbau Vocabularies are fixed with a size of X tokens [4, 11]. Words that are not part of vocabulary are represented as subwords and characters. Tokenizer takes the input sentence and will decide to keep every word as a whole word, split it into subwords (with a special representation of the first subword and subsequent subwords), or as a last resort decompose the word into individual characters.

2.2.2 Segment ID

BERT and BERTimbau are trained on and expect sentence pairs, using 1s and 0s to distinguish between the two sentences [4, 11]. That is, for each token in "tokenized text" we must specify which sentence it belongs to: sentence 0 (a series of 0s) or sentence 1 (a series of 1s). Single-sentence inputs only require a series of 1s for each token in the input sentence. To process two sentences, it must be assigned each word in the first sentence plus the '[SEP]' token a 0, and all tokens of the second sentence a 1.

2.2.3 Embeddings

With the divided tokens and segment IDs defined (each with a size of 512), the embeddings were collected through position 2 of each model output (BERT and BERTimbau). To use these embeddings, from each document and for each model, in our KNN classifier, we converted the tensor embeddings to numpy arrays and then concatenate them, for each document.

2.3 KNN

KNN (K-Nearest Neighbors) is an unsupervised method of classification based on closest examples in the feature space [5, 6]. KNN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until classification. The KNN method can be used to predict labels of any type. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors where k is a positive integer. KNN is analytically tractable, highly adaptive to local information, is easily implemented in parallel and uses the closest data points for estimation, therefore it can take full advantage of local information and form highly nonlinear, highly adaptive decision boundaries for each data point [6].

Predictions are often computed as a weighted average of deviations from neighbor means, as shown in Equation 1:

$$p_{a,i} = \vec{r_a} \frac{\sum (r_{b,i} - \vec{r_b}) \times w_{a,b}}{\sum w_{a,b}} \tag{1}$$

Where: prediction $p_{a,i}$ of item i for user a is an average of the set of deviations $(r_{b,i}\vec{(r_b)})$ from each neighbor's mean rating $\vec{(r_b)}$, weighted according to the similarity $w_{a,b}$ between the user a, and neighbour b. Other methods only consider the ratings themselves, rather than the deviations. However, all methods share the fact that they weigh the contribution of each neighbor according to the degree of similarity shared with the current user: similarity is central to this process.

$w_{a,b}$ can be calculated with Pearson correlation coefficient (PCC), as shown in Equation 2:

$$w_{a,b} = \frac{\sum_{i=1}^{N} (r_{a,i} - \vec{r_a})(r_{b,i} - \vec{r_b})}{\sqrt{\sum_{i=1}^{N} (r_{a,i} - \vec{r_a})^2 \sum_{i=1}^{N} (r_{b,i} - \vec{r_b})^2}} \tag{2}$$

Because it is instance-based, for each data point to be scored, the algorithm checks against the training table for the k nearest neighbor [6]. Since each data point is independent of the others, the execution of search and score can be conducted in parallel. The training samples are described by n-dimensional numeric attributes. The training samples are stored in an n-dimensional space. When a test sample (unknown class label) is given, the k-nearest neighbor classifier searches the k training samples which are closest to the unknown sample. The new instance is checked with the already available cases, based on distance assignment and classified using k value. More the instances are similar; least is the distance assigned and vice-versa.

Although KNN has many advantages, it has some disadvantages, like its high computation cost since it needs to compute the distance of each test instance to all training samples [6]. Also, it requires large memory proportional to the size of the training set and has a low accuracy rate in multidimensional data sets with irrelevant features. There is no thumb rule to determine value of parameter k (number of nearest neighbors). KNN classifier is k dependent since it predicts the outcome based on the value of k (number of nearest neighbors). Hence, for different values of k, the outcome may not be the same. KNN delays the process of modeling the training data until it is needed to classify the test samples.

3 Dataset

The “Querido Diário” project currently has around 56,000 documents, each containing hundreds to thousands of words, from official journals of 12 Brazilian capitals with complete temporal coverage – that is, the start date of the documents depends on when the document was made available by each city hall.

3.1 Data availability

The official journals on “Querido Diário” are available in three different ways:

1. Calls can be made to the content of the “Dear Diary” via API. However, for this alternative, journal text is not returned;

2. A copy of the Elasticsearch index of August 2021 can be made, provided that access is made available by the “Dear Diary” team. To use the index, you need to do an Elasticsearch snapshot restore;

3. The “Querido Diário” project also has in its repository on Github [8] links available to 182 official journals (178 without duplicates) on Google Drive.

Initially, we tried to collect the official journals through Elasticsearch. However, this task proved to be time-consuming. Because of the short period to finish this work, we chose to go with the last option, working with the 178 official journals available on Google Drive. We have provided links to these 182 documents in our repository [10].

The datasets of “Querido Diário” do not have any tags or labels provided. So we annotate and make available in our repository [10] City/State tags for all the documents from Google Drive and Section tags (detailed in the next subsection) for 20 documents from 8 different states. The links and tags are also shown in Appendix (Section 7).

We downloaded the documents using the **urllib** library. Then, we extracted the texts and available information from the PDF files with the **PyPDF** library.

Figure (a) shows the number of documents proportion for the 182 documents obtained through Google Drive from “Querido Diário”. The state with the most documents is Rio de Janeiro, with 104 official journals available. The state with the fewest documents is Alagoas, with only 1 official journal available. Figure (b) shows the means and standard deviations of the number of pages for documents from each observed state. The state with the most pages per document is Tocantins, with 96 pages. Rio de Janeiro also presented a high number of pages, and with more documents than Tocantins, with an average of 88 pages.

3.2 Official journal's sections

A municipal journal is categorized logically into a few sections [7]. These sections correspond to legal requirements for the publication of public acts. However, as cities do not follow a uniform standard for disclosing information, these sections do not have common titles, do not appear with the same structure and are not published in all official journals – after all, not always all public acts whose publication is required by law occur every day. Generally speaking, the mandatory sections are:

1. Normative Acts: Obligations or rights to citizens resulting from the activity of the municipal executive (or legislative) power;

2. Personnel Acts: Changes in the activities of municipal employees, contemplating entry into a public career, dismissal, granting temporary leave, etc;

3. Public Accounts: Cities debt obligations and budget request in respect of Fiscal Responsibility Law (Complementary Law No. 101, of 05/04/2000 [2]);

4. Public Hearings: Informs the population about public hearings regarding the activity of the municipal government and the implementation of local public policies;

5. Biddings: Reports on the main contracting instruments for private goods and services in Brazil by the public sector;

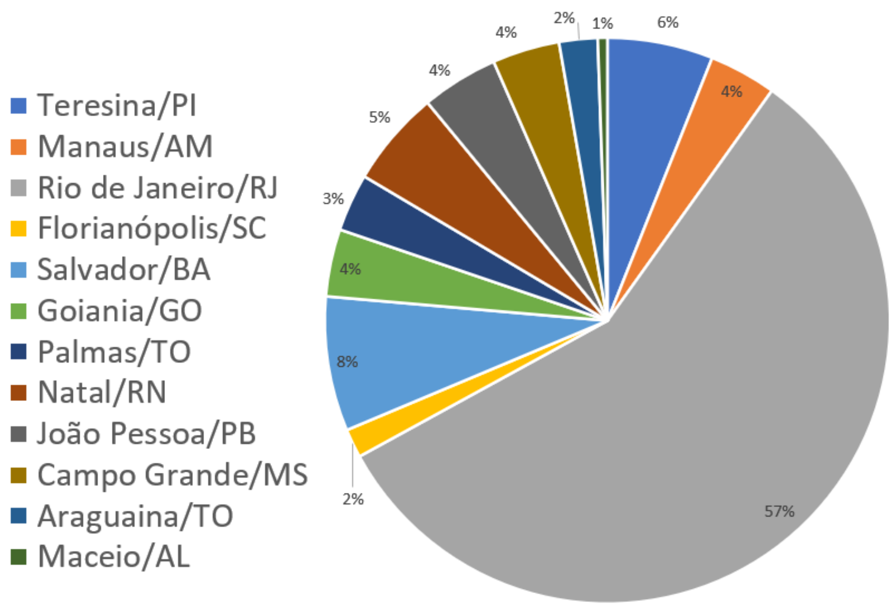
6. Disciplinary Proceedings: Reports of investigative and disciplinary proceedings on the conduct of public servants or contractors of the municipal executive power;

7. Tax Proceedings: Reports of judgments between the municipal tax authorities and taxpayers in tax administrative proceedings that affect the city's collection;

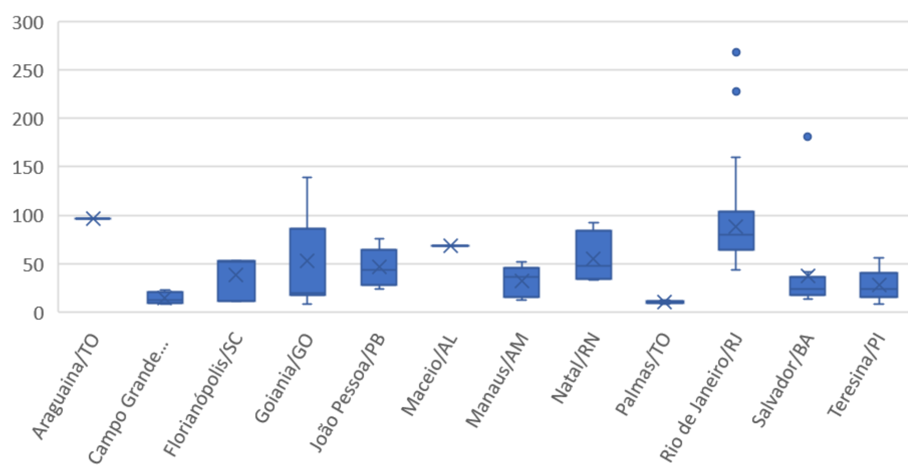
8. Municipal Councils: Reports on the inspection of public policies by civil society.

4 Experiments

Our experiments, following the methodology described on Chapter 2, have been done creating an array of embeddings produced by BERTimbau from each document; then we have generated a Pearson's correlation matrix to indicate the similarity distance among the files. This process can be seen in Figure 4, whereas the resulting matrix's rows and columns represent the ids of the documents. It



(a) Number of documents for each state



(b) Mean and standard deviation of the number of pages for each state

Figure 2: Graphs for Dataset analysis. (a) Proportion of the number of documents for each state available in Google Drive of the project "Querido Diário"; (b) Mean and standard deviation of the number of pages for each state documents available in Google Drive of the project "Querido Diário".

is notable that the main diagonal contains ones, representing the Pearson correlation of a document to itself being one, as expected.

In our experiments, we removed the official journals related to the city of Rio de Janeiro, due to the discrepancy in the absolute number of journals (over 50% of the documents, which would make the recommendation system recommend Rio's journals in an unbalanced frequency) and the discrepancy of the number of pages (Rio's are in general more than double the size of the other state's journals), as registered in the Chapter 3.

That exclusion resulted in 43 official journals that would go through BERTimbau and have their embedding layer's output saved in an array that would identify which city it belongs to.

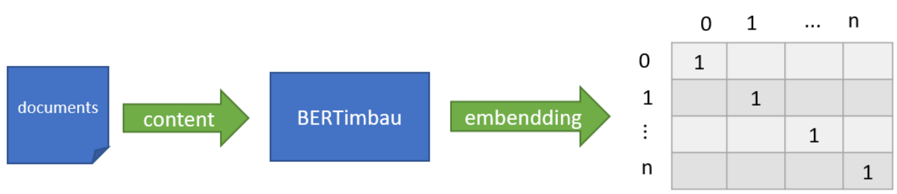


Figure 3: Matrix Creation process following the flowchart proposed on Chapter 2.

Using the created matrix we could infer the K neighbors near for each of those documents, according to Equation 1, and recommend the closest K documents for a given document D. With the implementation of Equation 1, we started to test different numbers of K neighbors and measure the average Pearson correlation among the documents (as shown in Equation 2). In order to obtain a comparison basis for the performance of our solution, we performed the same processes described in section 2, but using the BERT tokenizer and the BERT model to collect embeddings. The results can be seen in Table 1.

To exemplify the recommendation and to compare the two approaches (using BERTimbau and BERT), we have performed the KNN using K=1 comparing the id of the files and where they cam from. The results are shown in Table 2.

Observing the Table 2 it's notable that the recommendation on both algorithms favors states that are geological near if not in the same state, that could be possible because the dairies treat of similar topics or the writing similarities among those states showing the regional differences in Brazilian Portuguese.

5 Conclusion

In our work, we explored the use of embeddings acquired through the BERTimbau and BERT models to feed a KNN classifier with a document recommendation task for the "Querido Diário" project. Considering that the available datasets did not have any annotated data, the calculation of similarities and

Table 1: Average Pearson Correlation over K

K	**Average Correlation using BERTimbau**	**Average Correlation using BERT**
1	0.9358	0.9082
2	0.9324	0.9039
3	0.9298	0.9008
4	0.9272	0.8982
5	0.925	0.896
6	0.9231	0.8941
7	0.9215	0.8924
8	0.9199	0.8909
9	0.9183	0.8895
10	0.9169	0.8882
11	0.9155	0.887
12	0.9143	0.8858
13	0.9131	0.8847
14	0.9119	0.8836
15	0.9108	0.8826
16	0.9098	0.8815
17	0.9087	0.8805
18	0.9077	0.8795
19	0.9066	0.8786
20	0.9056	0.8776

Table 2: Recommendation using BERTimbau and BERT

id	Document State	BERTimbau Suggestion	BERTimbau's Suggestion State	BERTSuggestion	BERT's Suggestion State
0	AL	31	PB	28	PB
1	AM	37	TO	41	TO
2	AM	34	SC	8	BA
3	AM	15	BA	15	BA
4	AM	34	SC	17	BA
5	AM	38	TO	4	AM
6	AM	13	BA	3	AM
7	BA	18	BA	25	PB
8	BA	34	SC	0	AL
9	BA	26	PB	12	BA
10	BA	26	PB	17	BA
11	BA	6	AM	3	AM
12	BA	10	BA	32	PB
13	BA	34	SC	25	PB
14	BA	38	TO	38	TO
15	BA	3	AM	3	AM
16	BA	29	PB	38	TO
17	BA	34	SC	24	GO
18	BA	7	BA	25	PB
19	BA	15	BA	24	GO
20	GO	40	TO	33	SC
21	GO	41	TO	28	PB
22	GO	10	BA	36	TO
23	GO	41	TO	41	TO
24	GO	3	AM	17	BA
25	PB	13	BA	10	BA
26	PB	10	BA	10	BA
27	PB	34	SC	17	BA
28	PB	2	AM	21	GO
29	PB	30	PB	36	TO
30	PB	29	PB	29	PB
31	PB	21	GO	41	TO
32	PB	12	BA	12	BA
33	SC	10	BA	20	GO
34	SC	4	AM	17	BA
35	SC	42	TO	24	GO
36	TO	40	TO	41	TO
37	TO	14	BA	17	BA
38	TO	14	BA	4	AM
39	TO	10	BA	36	TO
40	TO	36	TO	36	TO
41	TO	21	GO	36	TO
42	TO	13	BA	35	SC

the use of KNN proved to be very efficient to return the K most similar documents, given a D document.

Comparing the average correlations for each model (BERTimbau and BERT), we noticed that the use of BERTimbau embeddings promoted higher average correlations than the cases in which BERT embeddings were used, for all tested K values. It was also noticed that the content recommendation favors regions close to each other geographically, which could indicate the regionalism of Brazilian Portuguese or the documents approach similar topics.

In conclusion, we have successfully built a recommendation system using "Querido Diário"'s official gazettes applying the knowledge acquired on IA - IA376.

6 Future Work

Among possible future works, we believe it is possible:

1. Use a machine with more memory, allowing to process all documents;
2. Get the Elastic Search image from "Querido Diário" and process all documents;
3. Optimize token extraction and conversion to embeddings, to execute it quickly;
4. Change how the embeddings are converted, increasing the window of 512 embeddings that were used due to the use of BERTimbau and BERT to avoid the need for concatenation of paragraphs in each document;
5. Investigate deeper the reason why BERTimbau and BERT recommend states close to each other;
6. Use the Section tags (available in our repository [10] and in Appendix (Section 7)) to increase semantic performance;
7. Use another classifier for the classification task, as long as the needs of labels for training and validation are observed.

References

[1] Open knowledge brasil (okbr). *Available at:* https://ok.org.br.

[2] Fernando Cardoso, Pedro Malan, and Martus Tavares. Lei complementar nº 101, de 4 de maio de 2000. *Available at:* http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm, 2000.

[3] Humberto Castelo Branco, Mem Sá, Zilmar Macedo, Juracy Magalhães, Octavio Bulhões, Juarez Távora, Ney Braga, Pedro Aleixo, Walter Barcellos, Eduardo Gomes, Mathias Silva, Paulo Martins, Mauro Thibau,

Roberto Campos, and Osvaldo Farias. Lei no 4.965, de 5 de maio de 1966. *Available at:* http://www.planalto.gov.br/ccivil_03/leis/1950-1969/l4965.htm, 1966.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[5] Neal Lathia, Stephen Hailes, and Licia Capra. knn cf: A temporal social network. In *ACM Conference on Recommender Systems*. ACM, 2008.

[6] M. NirmalaDevi, S. Appavu alias Balamurugan, and U. V. Swathi. An amalgam knn to predict diabetes mellitus. In *IEEE International Conference on Emerging Trends in Computing Communication and Nanotechnology*. IEEE, 2008.

[7] Open Knowledge Brazil (OKBR). Querido diário. *Available at:* https://queridodiario.ok.org.br, 2021.

[8] Open Knowledge Brazil (OKBR). Querido diário - github repository. *Available at:* https://github.com/okfn-brasil/querido-diario, 2021.

[9] Open Knowledge Brazil (OKBR). Querido diário - swagger ui. *Available at:* https://queridodiario.ok.org.br/api/docs#/default, 2021.

[10] Leonardo Rossi and Renan Yamaguti. Recomendação de documentos no domínio jurídico para o projeto querido diário - github repository. *Available at:* https://github.com/leolellisr/npl_recom_querido_diario.

[11] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. Bertimbau: Pretrained bert models for brazilian portuguese. In Ricardo Cerri and Ronaldo C. Prati, editors, *Intelligent Systems*, pages 403–417. Springer International Publishing, 2020.

7 Appendix

7.1 Links to documents and City/State tags

Tables 3, 4, 5 and 6 presents the links for the official journals available on the "Querido Diário" Google Drive and the City/State tags for 182 document.

7.2 Section tags

Here we share the Section tags for 20 official journals (from 8 diffcity and states) available on the "Querido Diário" Google Drive. The documents indexes follow the order of Tables 3, 4, 5 and 6. The Section tags were annotated according to the sections described in 3. Only the first 94 paragraphs of each document were considered, as only 2 documents had about 100 paragraphs. Tokens were defined with BERTimbau tokenizer.

Table 3: Links available on the "Querido Diário" Google Drive and annotated City/State tags for 182 documents (Part 1/4)

Index	Link	City/State tags
0	http://dom.pmt.pi.gov.br/admin/upload/DOM3089-19082021-ASSINADO.pdf	Teresina/PI
1	http://dom.manaus.am.gov.br/pdf/2014/marco/DOM%203379%2027.03.2014%20CAD%201.pdf	Manaus/AM
2	https://doweb.rio.rj.gov.br/portal/edicoes/download/2232	Rio de Janeiro/RJ
3	http://dom.manaus.am.gov.br/pdf/2017/janeiro/DOM%204043%2011.01.2017%20CAD%201.pdf	Manaus/AM
4	http://www.pmf.sc.gov.br/arquivos/diario/pdf/25_09_2013_20.54.32.f01bf9f1a8cbf65fb35be89e67d223e6.pdf	Florianópolis/SC
5	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2009/dezembro/DOM-5027-04-12-2009.pdf	Salvador/BA
6	http://dom.manaus.am.gov.br/pdf/2021/junho/DOM%205122%2018.06.2021%20CAD%201.pdf	Manaus/AM
7	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/1993/do_19931210_000001060.pdf	Goiania/GO
8	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2008/novembro/dom-4773-06-11-2008.pdf	Salvador/BA
9	http://diariooficial.palmas.to.gov.br/media/diario/877-31-10-2013.pdf	Palmas/TO
10	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2007/do_20070823_000004189.pdf	Goiania/GO
11	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20200522_e76670dcea11f8265ac4f6d5a939f5ef.pdf	Natal/RN
12	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2009/junho/DOM-4913-11-06-2009.pdf	Salvador/BA
13	https://doweb.rio.rj.gov.br/portal/edicoes/download/1237	Rio de Janeiro/RJ
14	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2011/04/2011_1267.pdf	João Pessoa/PB
15	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2021/05/2021_1788.pdf	João Pessoa/PB
16	https://doweb.rio.rj.gov.br/portal/edicoes/download/3758	Rio de Janeiro/RJ
17	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=7183&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
18	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20160406.pdf	Natal/RN
19	https://doweb.rio.rj.gov.br/portal/edicoes/download/2524	Rio de Janeiro/RJ
20	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2009/outubro/DOM-4991-09-10-2009.pdf	Salvador/BA
21	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2017/do_20170921_000006657.pdf	Goiania/GO
22	https://doweb.rio.rj.gov.br/portal/edicoes/download/1947	Rio de Janeiro/RJ
23	https://doweb.rio.rj.gov.br/portal/edicoes/download/3248	Rio de Janeiro/RJ
24	https://doweb.rio.rj.gov.br/portal/edicoes/download/3068	Rio de Janeiro/RJ
25	https://doweb.rio.rj.gov.br/portal/edicoes/download/3259	Rio de Janeiro/RJ
26	https://doweb.rio.rj.gov.br/portal/edicoes/download/2815	Rio de Janeiro/RJ
27	https://doweb.rio.rj.gov.br/portal/edicoes/download/4213	Rio de Janeiro/RJ
28	https://doweb.rio.rj.gov.br/portal/edicoes/download/4107	Rio de Janeiro/RJ
29	https://doweb.rio.rj.gov.br/portal/edicoes/download/3219	Rio de Janeiro/RJ
30	https://doweb.rio.rj.gov.br/portal/edicoes/download/4904	Rio de Janeiro/RJ
31	https://doweb.rio.rj.gov.br/portal/edicoes/download/3511	Rio de Janeiro/RJ
32	http://dom.pmt.pi.gov.br/admin/upload/ANEXO_DOM1195-1-28122007.pdf	Teresina/PI
33	https://doweb.rio.rj.gov.br/portal/edicoes/download/2877	Rio de Janeiro/RJ
34	https://doweb.rio.rj.gov.br/portal/edicoes/download/972	Rio de Janeiro/RJ
35	https://doweb.rio.rj.gov.br/portal/edicoes/download/271	Rio de Janeiro/RJ
36	https://doweb.rio.rj.gov.br/portal/edicoes/download/1034	Rio de Janeiro/RJ
37	https://doweb.rio.rj.gov.br/portal/edicoes/download/1337	Rio de Janeiro/RJ
38	https://doweb.rio.rj.gov.br/portal/edicoes/download/2503	Rio de Janeiro/RJ
39	https://doweb.rio.rj.gov.br/portal/edicoes/download/4630	Rio de Janeiro/RJ
40	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2019/do_20191003_000007153.pdf	Goiania/GO
41	https://doweb.rio.rj.gov.br/portal/edicoes/download/4063	Rio de Janeiro/RJ
42	https://doweb.rio.rj.gov.br/portal/edicoes/download/4226	Rio de Janeiro/RJ
43	https://doweb.rio.rj.gov.br/portal/edicoes/download/4632	Rio de Janeiro/RJ
44	https://doweb.rio.rj.gov.br/portal/edicoes/download/4642	Rio de Janeiro/RJ
45	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=7140&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
46	https://doweb.rio.rj.gov.br/portal/edicoes/download/4714	Rio de Janeiro/RJ
47	https://doweb.rio.rj.gov.br/portal/edicoes/download/2953	Rio de Janeiro/RJ
48	https://doweb.rio.rj.gov.br/portal/edicoes/download/4592	Rio de Janeiro/RJ
49	https://doweb.rio.rj.gov.br/portal/edicoes/download/3821	Rio de Janeiro/RJ
50	https://doweb.rio.rj.gov.br/portal/edicoes/download/4598	Rio de Janeiro/RJ

Table 4: Links available on the "Querido Diário" Google Drive and annotated City/State tags for 182 documents (Part 2/4)

Index	Link	City/State tags
51	https://doweb.rio.rj.gov.br/portal/edicoes/download/2936	Rio de Janeiro/RJ
52	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2019/dezembro/dom-7538-27-12-2019.pdf	Salvador/BA
53	https://doweb.rio.rj.gov.br/portal/edicoes/download/4848	Rio de Janeiro/RJ
54	https://doweb.rio.rj.gov.br/portal/edicoes/download/4085	Rio de Janeiro/RJ
55	https://doweb.rio.rj.gov.br/portal/edicoes/download/4126	Rio de Janeiro/RJ
56	https://doweb.rio.rj.gov.br/portal/edicoes/download/4903	Rio de Janeiro/RJ
57	https://doweb.rio.rj.gov.br/portal/edicoes/download/2593	Rio de Janeiro/RJ
58	https://doweb.rio.rj.gov.br/portal/edicoes/download/1804	Rio de Janeiro/RJ
59	https://doweb.rio.rj.gov.br/portal/edicoes/download/356	Rio de Janeiro/RJ
60	https://doweb.rio.rj.gov.br/portal/edicoes/download/2184	Rio de Janeiro/RJ
61	https://doweb.rio.rj.gov.br/portal/edicoes/download/4855	Rio de Janeiro/RJ
62	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=5396&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
63	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20180201_22d614bc09a3125710f95985323a71b6.pdf	Natal/RN
64	http://dom.manaus.am.gov.br/pdf/2019/abril/DOM%204579%2016.04.2019%20CAD%201.pdf	Manaus/AM
65	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2019/novembro/dom-7505-15-11-2019.pdf	Salvador/BA
66	https://doweb.rio.rj.gov.br/portal/edicoes/download/4302	Rio de Janeiro/RJ
67	https://doweb.rio.rj.gov.br/portal/edicoes/download/3095	Rio de Janeiro/RJ
68	http://dom.pmt.pi.gov.br/admin/upload/DOM1693-10122014.pdf	Teresina/PI
69	https://doweb.rio.rj.gov.br/portal/edicoes/download/1230	Rio de Janeiro/RJ
70	https://doweb.rio.rj.gov.br/portal/edicoes/download/473	Rio de Janeiro/RJ
71	https://doweb.rio.rj.gov.br/portal/edicoes/download/2252	Rio de Janeiro/RJ
72	https://doweb.rio.rj.gov.br/portal/edicoes/download/1245	Rio de Janeiro/RJ
73	http://dom.pmt.pi.gov.br/admin/upload/DOM1122-1-13102006.pdf	Teresina/PI
74	https://doweb.rio.rj.gov.br/portal/edicoes/download/4408	Rio de Janeiro/RJ
75	https://doweb.rio.rj.gov.br/portal/edicoes/download/524	Rio de Janeiro/RJ
76	https://doweb.rio.rj.gov.br/portal/edicoes/download/1046	Rio de Janeiro/RJ
77	http://dom.manaus.am.gov.br/pdf/2013/dezembro/DOM%203319%2026.12.2013%20CAD%201.pdf	Manaus/AM
78	https://doweb.rio.rj.gov.br/portal/edicoes/download/3063	Rio de Janeiro/RJ
79	https://doweb.rio.rj.gov.br/portal/edicoes/download/1084	Rio de Janeiro/RJ
80	https://doweb.rio.rj.gov.br/portal/edicoes/download/2548	Rio de Janeiro/RJ
81	https://doweb.rio.rj.gov.br/portal/edicoes/download/3076	Rio de Janeiro/RJ
82	https://doweb.rio.rj.gov.br/portal/edicoes/download/3833	Rio de Janeiro/RJ
83	https://doweb.rio.rj.gov.br/portal/edicoes/download/5075	Araguaina/TO
84	https://doweb.rio.rj.gov.br/portal/edicoes/download/5070	Rio de Janeiro/RJ
85	https://diariooficial.araguaina.to.gov.br/Arquivo/DiarioOficial/pdf/2015.pdf	Rio de Janeiro/RJ
86	https://doweb.rio.rj.gov.br/portal/edicoes/download/4940	Rio de Janeiro/RJ
87	https://doweb.rio.rj.gov.br/portal/edicoes/download/3114	Rio de Janeiro/RJ
88	https://doweb.rio.rj.gov.br/portal/edicoes/download/3261	Rio de Janeiro/RJ
89	https://doweb.rio.rj.gov.br/portal/edicoes/download/726	Rio de Janeiro/RJ
90	https://doweb.rio.rj.gov.br/portal/edicoes/download/1769	Rio de Janeiro/RJ
91	https://doweb.rio.rj.gov.br/portal/edicoes/download/1612	Rio de Janeiro/RJ
92	https://doweb.rio.rj.gov.br/portal/edicoes/download/789	Rio de Janeiro/RJ
93	https://doweb.rio.rj.gov.br/portal/edicoes/download/2736	Rio de Janeiro/RJ
94	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20141229.pdf	Natal/RN
95	https://doweb.rio.rj.gov.br/portal/edicoes/download/2484	Rio de Janeiro/RJ
96	https://doweb.rio.rj.gov.br/portal/edicoes/download/2409	Rio de Janeiro/RJ
97	https://doweb.rio.rj.gov.br/portal/edicoes/download/2538	Rio de Janeiro/RJ
98	https://doweb.rio.rj.gov.br/portal/edicoes/download/3169	Rio de Janeiro/RJ
99	https://doweb.rio.rj.gov.br/portal/edicoes/download/3509	Rio de Janeiro/RJ
100	https://doweb.rio.rj.gov.br/portal/edicoes/download/1209	Rio de Janeiro/RJ

Table 5: Links available on the "Querido Diário" Google Drive and annotated City/State tags for 182 documents (Part 3/4)

Index	Link	City/State tags
101	https://doweb.rio.rj.gov.br/portal/edicoes/download/3263	Rio de Janeiro/RJ
102	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2017/10/2017_1603.pdf	João Pessoa/PB
103	https://doweb.rio.rj.gov.br/portal/edicoes/download/2617	Rio de Janeiro/RJ
104	https://doweb.rio.rj.gov.br/portal/edicoes/download/2963	Rio de Janeiro/RJ
105	https://doweb.rio.rj.gov.br/portal/edicoes/download/2482	Rio de Janeiro/RJ
106	https://doweb.rio.rj.gov.br/portal/edicoes/download/1233	Rio de Janeiro/RJ
107	https://doweb.rio.rj.gov.br/portal/edicoes/download/4688	Rio de Janeiro/RJ
108	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20120802.pdf	Natal/RN
109	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20120605.pdf	Natal/RN
110	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2019/fevereiro/dom-7306-13-02-2019.pdf	Salvador/BA
111	https://doweb.rio.rj.gov.br/portal/edicoes/download/1004	Rio de Janeiro/RJ
112	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20101028.pdf	Natal/RN
113	http://diariooficial.palmas.to.gov.br/media/diario/1782-26-6-2017-19-40-51.pdf	Palmas/TO
114	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2011/09/2009_1175.pdf	João Pessoa/PB
115	https://doweb.rio.rj.gov.br/portal/edicoes/download/4192	Rio de Janeiro/RJ
116	https://doweb.rio.rj.gov.br/portal/edicoes/download/288	Rio de Janeiro/RJ
117	https://doweb.rio.rj.gov.br/portal/edicoes/download/1553	Rio de Janeiro/RJ
118	https://doweb.rio.rj.gov.br/portal/edicoes/download/1029	Rio de Janeiro/RJ
119	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2018/dezembro/dom-7271-27-12-2018.pdf	Salvador/BA
120	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2018/janeiro/dom-7018-11-01-2018.pdf	Salvador/BA
121	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2016/setembro/dom-6679-17-09-2016.pdf	Salvador/BA
122	https://doweb.rio.rj.gov.br/portal/edicoes/download/3078	Rio de Janeiro/RJ
123	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2017/do_20171204_000006704.pdf	Goiania/GO
124	http://diariooficial.palmas.to.gov.br/media/diario/1314-5-8-2015-19-48-17.pdf	Palmas/TO
125	http://dom.pmt.pi.gov.br/admin/upload/DOM1383-04022011.pdf-04022011.pdf-04022011	Teresina/PI
126	https://doweb.rio.rj.gov.br/portal/edicoes/download/309	Rio de Janeiro/RJ
127	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2010/do_20100514_000004861.pdf	Goiania/GO
128	https://doweb.rio.rj.gov.br/portal/edicoes/download/352	Rio de Janeiro/RJ
129	https://doweb.rio.rj.gov.br/portal/edicoes/download/2412	Rio de Janeiro/RJ
130	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=7194&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
131	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=3386&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
132	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2017/10/2017_1601.pdf	João Pessoa/PB
133	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2001/outubro/DOM-3082-31-10-2001.pdf	Salvador/BA
134	https://doweb.rio.rj.gov.br/portal/edicoes/download/3481	Rio de Janeiro/RJ
135	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2002/maio/DOM-3211-17-05-2002.pdf	Salvador/BA
136	http://www.maceio.al.gov.br/wp-content/uploads/2017/03/pdf/2017/03/Diario_Oficial_16_03_17_PDF.pdf	Maceio/AL
137	https://doweb.rio.rj.gov.br/portal/edicoes/download/2681	Rio de Janeiro/RJ
138	https://doweb.rio.rj.gov.br/portal/edicoes/download/2288	Rio de Janeiro/RJ
139	https://doweb.rio.rj.gov.br/portal/edicoes/download/3503	Rio de Janeiro/RJ
140	https://doweb.rio.rj.gov.br/portal/edicoes/download/3682	Rio de Janeiro/RJ
141	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2019/10/2019_1707.pdf	João Pessoa/PB
142	https://doweb.rio.rj.gov.br/portal/edicoes/download/3291	Rio de Janeiro/RJ
143	https://doweb.rio.rj.gov.br/portal/edicoes/download/1709	Rio de Janeiro/RJ
144	https://doweb.rio.rj.gov.br/portal/edicoes/download/3171	Rio de Janeiro/RJ
145	http://www.pmf.sc.gov.br/arquivos/diario/pdf/29_01_2018_18.32.44.f3360bad8463c10b99f92433f960a704.pdf	Florianópolis/SC
146	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2019/11/2019_1710.pdf	João Pessoa/PB
147	https://doweb.rio.rj.gov.br/portal/edicoes/download/4013	Rio de Janeiro/RJ
148	https://doweb.rio.rj.gov.br/portal/edicoes/download/1261	Rio de Janeiro/RJ
149	https://doweb.rio.rj.gov.br/portal/edicoes/download/2858	Rio de Janeiro/RJ
150	https://doweb.rio.rj.gov.br/portal/edicoes/download/3725	Rio de Janeiro/RJ

Table 6: Links available on the “Querido Diário” Google Drive and annotated City/State tags for 182 documents (Part 4/4)

Index	Link	City/State tags
151	https://doweb.rio.rj.gov.br/portal/edicoes/download/2530	Rio de Janeiro/RJ
152	https://doweb.rio.rj.gov.br/portal/edicoes/download/4144	Rio de Janeiro/RJ
153	https://doweb.rio.rj.gov.br/portal/edicoes/download/1045	Rio de Janeiro/RJ
154	http://diariooficial.palmas.to.gov.br/media/diario/2544-31-7-2020-20-18-9.pdf	Palmas/TO
155	https://diariooficial.araguaina.to.gov.br/Arquivo/DiarioOficial/pdf/2168.pdf	Araguaina/TO
156	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=8062&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
157	https://diariooficial.araguaina.to.gov.br/Arquivo/DiarioOficial/pdf/1422.pdf	Araguaina/TO
158	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20190312_be4ee7db98c802de35ab0c58b725010e.pdf	Natal/RN
159	https://doweb.rio.rj.gov.br/portal/edicoes/download/682	Rio de Janeiro/RJ
160	https://doweb.rio.rj.gov.br/portal/edicoes/download/3191	Rio de Janeiro/RJ
161	http://dom.pmt.pi.gov.br/admin/upload/DOM2300-13062018-ASSINADO.pdf	Teresina/PI
162	http://portal.capital.ms.gov.br/egov/downloadFile.php?id=4884&fileField=arquivo_dia_ofi&table=diario_oficial&key=id_dia_ofi&sigla_sec=diogrande	Campo Grande/MS
163	http://dom.pmt.pi.gov.br/admin/upload/DOM2496-04042019-ASSINADO.pdf	Teresina/PI
164	https://doweb.rio.rj.gov.br/portal/edicoes/download/2910	Rio de Janeiro/RJ
165	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20190110_c22a1ca1a785d1394940a328393fd8f9.pdf	Natal/RN
166	http://dom.pmt.pi.gov.br/admin/upload/DOM2653-21112019-ASSINADO.pdf	Teresina/PI
167	http://antigo.joaopessoa.pb.gov.br/portal/wp-content/uploads/2012/01/2011_1302.pdf	João Pessoa/PB
168	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2014/maio/dom-6099-22-05-2014.pdf	Salvador/BA
169	http://dom.pmt.pi.gov.br/admin/upload/DOM2656-26112019-ASSINADO.pdf	Teresina/PI
170	http://dom.pmt.pi.gov.br/admin/upload/DOM2673-19122019-ASSINADO.pdf	Teresina/PI
171	http://dom.pmt.pi.gov.br/admin/upload/DOM1237-1-05092008.pdf	Teresina/PI
172	https://diariooficial.araguaina.to.gov.br/Arquivo/DiarioOficial/pdf/1527.pdf	Araguaina/TO
173	http://dom.manaus.am.gov.br/pdf/2019/outubro/DOM%204707%2022.10.2019%20CAD%201.pdf	Manaus/AM
174	http://dom.manaus.am.gov.br/pdf/2007/novembro/dom20071849cad1.pdf	Manaus/AM
175	http://diariooficial.palmas.to.gov.br/media/diario/2673-9-2-2021-20-17-33.pdf	Palmas/TO
176	http://www.pmf.sc.gov.br/arquivos/diario/pdf/29_06_2018_17.55.09.d2e337b68e2e7409094954c1153687c2.pdf	Florianópolis/SC
177	https://doweb.rio.rj.gov.br/portal/edicoes/download/868	Rio de Janeiro/RJ
178	http://www.dom.salvador.ba.gov.br/images/stories/pdf/2008/novembro/dom-4773-06-11-2008.pdf	Salvador/BA
179	http://diariooficial.palmas.to.gov.br/media/diario/877-31-10-2013.pdf	Palmas/TO
180	http://www.goiania.go.gov.br/Download/legislacao/diariooficial/2007/do_20070823_000004189.pdf	Goiania/GO
181	http://portal.natal.rn.gov.br/_anexos/publicacao/dom/dom_20200522_e76670dcea11f8265ac4f6d5a939f5ef.pdf	Natal/RN

Table 7: Section tags for 20 documents (Part 1/2)

Index document	0	1	2	3	4	5	7	8	9	10	12	14	15	20	32	45	50	162	175	176
Paragraph	Section tag																			
0	0	0	0	1	6	0	0	0	0	1	0	1	0	0	1	0	0	6	0	6
1	0	0	0	1	6	0	6	0	0	6	0	1	0	0	1	0	0	6	1	6
2	0	0	0	1	6	0	6	0	4	6	0	1	0	0	1	0	1	6	1	1
3	0	1	0	1	6	0	2	6	4	6	2	1	0	0	1	0	1	6	6	1
4	0	1	0	1	6	6	1	6	4	6	2	1	0	1	1	0	1	6	-	1
5	0	1	0	1	6	6	1	6	4	0	2	1	0	1	1	0	1	6	-	1
6	0	2	0	1	0	6	1	6	4	0	2	1	0	6	0	0	1	4	-	1
7	0	2	0	1	1	6	1	6	4	0	2	1	0	6	0	1	0	0	-	2
8	1	2	0	1	1	6	1	6	4	0	2	1	0	6	0	1	0	0	-	2
9	1	1	0	1	1	6	4	6	4	-	0	1	0	6	0	1	1	0	-	1
10	1	1	2	1	1	6	4	6	4	-	0	1	0	1	0	1	1	0	-	2
11	1	1	0	1	1	6	-	6	4	-	1	1	0	1	0	1	0	0	-	4
12	1	1	0	1	1	6	-	6	4	-	1	1	0	1	0	1	0	0	-	0
13	2	1	0	1	1	6	-	6	4	-	0	1	0	1	0	1	0	0	-	0
14	2	1	0	1	1	6	-	6	4	-	0	1	0	0	0	1	0	0	-	0
15	2	1	0	1	1	6	-	6	4	-	4	1	0	0	0	1	0	0	-	-
16	2	1	0	1	1	1	-	6	4	-	4	1	0	0	0	1	6	0	-	-
17	2	1	1	1	1	1	-	1	-	-	4	1	1	0	0	1	6	0	-	-
18	2	1	1	6	1	2	-	1	-	-	0	1	1	0	0	1	6	0	-	-
19	2	1	6	6	1	2	-	6	-	-	0	1	1	0	0	1	6	0	-	-
20	2	1	6	6	1	0	-	1	-	-	0	1	1	0	0	5	6	0	-	-
21	2	0	6	6	1	0	-	1	-	-	0	1	1	0	0	5	6	0	-	-
22	2	0	6	1	1	1	-	1	-	-	0	1	1	0	0	4	6	0	-	-
23	2	2	6	2	1	1	-	1	-	-	0	1	1	0	0	4	6	0	-	-
24	2	2	6	1	4	1	-	1	-	-	0	1	1	0	0	4	6	0	-	-
25	2	2	6	1	4	1	-	1	-	-	-	1	1	0	0	0	6	0	-	-
26	2	2	6	1	4	1	-	4	-	-	-	1	1	4	0	0	6	0	-	-
27	2	2	6	0	2	1	-	4	-	-	-	1	1	4	0	0	6	4	-	-
28	2	0	6	0	2	1	-	4	-	-	-	1	1	4	0	0	6	4	-	-
29	2	0	6	0	2	0	-	4	-	-	-	1	1	4	0	0	6	1	-	-
30	2	0	6	0	2	0	-	0	-	-	-	1	1	4	0	0	6	1	-	-
31	-	0	6	0	2	0	-	0	-	-	-	1	1	4	0	0	6	-	-	-
32	-	0	6	0	2	0	-	0	-	-	-	1	1	0	0	0	5	-	-	-
33	-	0	6	0	2	0	-	0	-	-	-	1	1	0	0	0	0	-	-	-
34	-	0	6	0	2	0	-	0	-	-	-	1	1	0	0	0	2	-	-	-
35	-	0	2	0	2	0	-	0	-	-	-	1	1	0	0	0	2	-	-	-
36	-	0	2	0	2	0	-	0	-	-	-	1	1	0	0	0	2	-	-	-
37	-	0	2	0	2	0	-	0	-	-	-	6	1	0	0	1	2	-	-	-
38	-	0	2	0	2	0	-	0	-	-	-	6	1	0	0	1	0	-	-	-
39	-	0	2	0	2	0	-	0	-	-	-	4	1	0	0	-	0	-	-	-
40	-	0	2	0	2	0	-	0	-	-	-	4	0	0	0	-	0	-	-	-
41	-	0	2	0	2	0	-	0	-	-	-	4	0	0	0	-	0	-	-	-
42	-	0	2	0	2	0	-	0	-	-	-	4	0	-	0	-	1	-	-	-
43	-	0	2	0	2	0	-	0	-	-	-	4	0	-	0	-	1	-	-	-
44	-	0	2	0	2	0	-	-	-	-	-	4	0	-	0	-	0	-	-	-
45	-	0	2	0	-	0	-	-	-	-	-	4	0	-	0	-	0	-	-	-
46	-	0	2	0	-	0	-	-	-	-	-	4	0	-	0	-	0	-	-	-
47	-	0	4	0	-	0	-	-	-	-	-	4	-	-	0	-	0	-	-	-
48	-	0	4	0	-	0	-	-	-	-	-	4	-	-	0	-	0	-	-	-
49	-	0	2	0	-	0	-	-	-	-	-	4	-	-	0	-	0	-	-	-

Table 8: Section tags for 20 documents (Part 2/2)

Index document	0	1	2	3	4	5	7	8	9	10	12	14	15	20	32	45	50	162	175	176
Paragraph	Section tag																			
50	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
51	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
52	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
53	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
54	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
55	-	0	2	0	-	4	-	-	-	-	-	4	-	-	0	-	0	-	-	-
56	-	0	2	0	-	4	-	-	-	-	-	4	-	-	4	-	0	-	-	-
57	-	0	2	0	-	0	-	-	-	-	-	4	-	-	4	-	0	-	-	-
58	-	0	2	0	-	0	-	-	-	-	-	4	-	-	4	-	0	-	-	-
59	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
60	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
61	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
62	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
63	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
64	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
65	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
66	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
67	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
68	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
69	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
70	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
71	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
72	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
73	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
74	-	0	1	0	-	0	-	-	-	-	-	4	-	-	-	-	0	-	-	-
75	-	0	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
76	-	0	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
77	-	0	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
78	-	0	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
79	-	0	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
80	-	2	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
81	-	2	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
82	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
83	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
84	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
85	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
86	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
87	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
88	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
89	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	0	-	-	-
90	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	1	-	-	-
91	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	1	-	-	-
92	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	1	-	-	-
93	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	1	-	-	-
94	-	-	1	0	-	-	-	-	-	-	-	4	-	-	-	-	6	-	-	-