


## International Conference Artificial Intelligence in Engineering and Science

AIES 2022: **Artificial Intelligence in Models, Methods and Applications**  
pp 219–229

[Home](#) > [Artificial Intelligence in Models, Methods and Applications](#) > Conference paper

# TableProcessor: The Tool for the Analysis and the Interpretation of Web Tables to Create the Geo Knowledge Base of Kazakhstan

[Vladimir Barakhnin](#), [Madina Mansurova](#) , [Irina Grigorieva](#), [Olga Kozhemyakina](#) & [Assel Ospan](#)

Conference paper | [First Online: 25 April 2023](#)

**69** Accesses

Part of the [Studies in Systems, Decision and Control](#) book series (SSDC,volume 457)

## Abstract

---

The data retrieval from the tables is actual task, since the number of publicly available high-quality tables on the Internet containing useful relational data reaches hundreds of millions. The search engines usually ignore the basic semantics of such structures when indexing and do not work well with tabular data. The web tables do not adhere to any

single presentation scheme, what, of course, is a minus for the task. This article proposes the method for disclosing the semantics of tables, followed by a description in the form of a knowledge base suitable for automatic analysis and based on the heuristic selection of a candidate object for the main thematic entity with the subsequent calculation of the suitability assessment of each selected entity for the objects in the lines of the table and similarity assessment between the pairs of candidate objects to determine the best likelihood assessments of entities by line. The tool was developed for interpreting semi-structured tables from the open website of the Bureau of National Statistics of the Republic of Kazakhstan. The files JSON were obtained, for which the title, the main column of objects, the attributes, their values and the text after the tables are automatically detected. These files are ready for the usage in GIS to display the attribute data of the regions of Kazakhstan. The TableProcessor tool was developed for interpreting and analyzing web tables, based on the knowledge base on geo-objects of Kazakhstan will be formed, which will be supplemented with new knowledge.

## Keywords

**Semantic analysis      Geo-objects of Kazakhstan**

---

This is a preview of subscription content, [access via your institution](#).

---

▼ Chapter

EUR 29.95

Price includes VAT (Kazakhstan)

- DOI: 10.1007/978-3-031-22938-1\_15
- Chapter length: 11 pages
- Instant PDF download
- Readable on all devices
- Own it forever
- Exclusive offer for individuals only
- Tax calculation will be finalised during checkout

Buy Chapter

> eBook

EUR 181.89

> Hardcover Book

EUR 219.99

[Learn about institutional subscriptions](#)

## References

---

1. Barakhnin, V.B., Fedotov, A.M.: Building models of documentary and factographic retrieval in digital libraries. *Autom. Doc. Math. Linguist.* **48**(6), 296–304 (2014)
  2. Wittgenstein, L.: Logisch-philosophische abhandlung. *Annalen der Naturphilosophie*, vol. XIV(3/4). Verlag Unesma, S.185–262 (1921)
  3. Chen, P.P.: The entity-relational model. Toward a unified view of data. *ACM TODS* **1**, 9–36 (1976)
-

4. Cafarella, M., Halevy, A., Lee, H., Madhavan, J., Yu, C., Wang, Z.D., Wu, E.: Ten years of webtables. *PVLDB* **11**(12), 2140–2149 (2018).  
<https://doi.org/10.14778/3229863.3240492>

---

5. Mulwad, V., Finin, T., Joshi, A.: Semantic message passing for generating linked data from tables. In: *International Semantic Web Conference, Lecture Notes in Computer Science*, pp. 363–378. Springer, Berlin, Heidelberg (2013).  
[https://doi.org/10.1007/978-3-642-41335-3\\_23](https://doi.org/10.1007/978-3-642-41335-3_23)

---

6. Yakout, M., Ganjam, K., Chakrabarti, K., Chaudhuri, S.: InfoGather: entity augmentation and attribute discovery by holistic matching with web tables. In: *Proceedings of SIGMOD*, pp. 97–108 (2012)

---

7. Hayes, P.: *RDF Semantics*. W3C Recommendation (2004).  
<http://www.w3.org/TR/rdf-mt/>

---

8. Limaye, G., Sarawagi, S., Chakrabarti, S.: Annotating and searching web tables using entities types and relationships. *PVLDB* **3**(1–2), 1338–1347 (2010)

---

9. Lehmborg, O., Ritzke, D., Meusel, R., Bizer, C.: A large public corpus of web tables containing

time and context metadata. In: Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion, pp. 75–76 (2016)

---

10. Zhang, Sh., Balog, K.: Ad hoc table retrieval using semantic similarity. In: Proceedings of the 2018 World Wide Web Conference, WWW '18, pp. 1553–1562 (2018)

---

11. Chen, X., Chiticariu, L., Danilevsky, M., Evfimievski, A., Sen, P.: A rectangle mining method for understanding the semantics of financial tables. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pp. 268–273 (2017). <https://doi.org/10.1109/ICDAR.2017.52>

---

12. Kunihiro, T., Masafumi, O., Shinji, N., Takeshi, O.: An efficient probabilistic approach for semantically annotating tables. In: 33rd AAAI Conference on Artificial Intelligence, AAAI 2019 (2019)

---

13. Zhang, Z.: Effective and efficient semantic table interpretation using tableminer+. Semantic Web **8**(6), 921–957 (2017)

---

14. Nugumanova, A., Akhmed-Zaki, D., Mansurova, M., Baiburin, Y., Maulit, A.: NMF-based approach to automatic term extraction. *Expert Syst. Appl.* **199**, 117179 (2022)

---

15. Mansurova, M., Barakhnin, V., Kyrgyzbayeva, M., Kadyrbek, N.: Named entity extraction model based on the random walk method SIST 2021. In: 2021 IEEE International Conference on Smart Information Systems and Technologies, 9465992 (2021)

---

16. Zhang, Sh., Balog K.: Web table extraction, retrieval and augmentation. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (2019)

---

17. Hignette, G., Buche, P., Dizie-Barthélemy, J., Haemmerlé, O.: Fuzzy annotation of web data tables driven by a domain ontology. In: Proceedings of the Z. Zhang / Effective and Efficient Semantic Table Interpretation using TableMiner+ 31 6th European Semantic Web Conference on The Semantic Web: Research and Applications, ESWC'2009, pages 638– 653, Berlin, Heidelberg (2009). Springer-Verlag.  
[https://doi.org/10.1007/978-3-642-02121-3\\_47](https://doi.org/10.1007/978-3-642-02121-3_47)

---

18. Buche, P., Dizie-Barthélemy, J., Ibanescu, L., Soler, L.: Fuzzy web data tables integration guided by an ontological and terminological resource. *IEEE Trans. Knowl. Data Eng.* **25**(4), 805–819 (2013).  
<https://doi.org/10.1109/TKDE.2011.245>

---
19. Cremaschi, M., Paoli, F.D., Rula, A., Blerina, S.: A fully automated approach to a complete Semantic Table Interpretation. *Futur. Gener. Comput. Syst.* **112**, 478–500 (2020)

---
20. Chabot, Y., Labbe, T., Liu, J., Troncy, R.: An end-to-end context-free tabular data semantic annotation system. In: Proceedings of the Semantic Web Challenge on Tabular Data To Knowledge Graph Matching Co-Located with the 18th International Semantic Web Conference, ISWC 2019, CEUR-WS.org (2019)

---
21. Graphical Model  
<https://www.sciencedirect.com/topics/computer-science/graphical-model>

---
22. Chen, J., Jiménez-Ruiz, E., Horrocks, I., Sutton, C.C.: Embedding the semantics of web tables for column type prediction. In: 33rd AAAI Conference on Artificial Intelligence, AAAI 2019 (2019)

---

23. Chen, X., Chitikariu, L., Danilevsky, M., Evfimievski, A., Sen, P.: Box analysis method for understanding the semantics of financial tables. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pp. 268–273 (2017).  
<https://doi.org/10.1109/ICDAR.2017.52>.  
<https://www.cs.cornell.edu/~xlchen/resources/papers/TableExtraction.pdf>

---

24. Chen, Zh., Cafarella, M.: Automatic web spreadsheet data extraction. In: SSW '13: Proceedings of the 3rd International Workshop on Semantic Search Over the Web August 2013, Article No.: 1, pp. 1–8.  
<https://doi.org/10.1145/2509908.2509909>

---

25. Barowy, D.W., Gulwani, S., Hart, T., Zorn, B.: FlashRelate: extracting relational data from semi-structured spreadsheets using examples. In: PLDI '15: Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation, pp. 218–228 (2015).  
<https://doi.org/10.1145/2737924.2737952>

---

26. Benno, K., Boncz PA, Jacopo, U.: Extracting Novel Facts from Tables for Knowledge Graph Completion (Extended version), CoRR,



abs/1907.00083 (2019).

<http://arxiv.org/abs/1907.00083>

---

27. Pearl, J.: Probabilistic reasoning in intelligent systems - networks of plausible inference. Morgan Kaufmann Publishers Inc. (1989)

---

28. Varish, M., Tim, F., Anupam, J.: Semantic message passing for generating linked data from tables. In: International Semantic Web Conference, Lecture Notes in Computer Science, pages 363–378. Springer Berlin Heidelberg (2013).

[https://doi.org/10.1007/978-3-642-41335-3\\_23](https://doi.org/10.1007/978-3-642-41335-3_23)

---

29. Nugumanova, A.B., Apayev, K.S., Baiburin, Y.M., Mansurova, M., Ospan, A.G.: Qurma: the pipeline for extracting tables to replenish the knowledge base. Bulletin of the KazNU. Mathematics, Mechanics, Computer Science series, [S.l.], vol. 114, no. 2 (2022). ISSN 2617-4871.

<https://bm.kaznu.kz/index.php/kaznu/article/view/1086>

---

## Acknowledgements

---

This work was carried out and sponsored within the framework of the scientific project AP09261344 "Development of methods for automatic extraction

of spatial objects from heterogeneous sources for information support of geographic information systems”.

## Author information

---

Authors and Affiliations

**Federal Research Center for Information and Computing Technologies, Novosibirsk, Russia**

Vladimir Barakhnin, Irina Grigorieva & Olga Kozhemyakina

**Al Farabi Kazakh National University, Almaty, Kazakhstan**

Madina Mansurova & Assel Ospan

Corresponding author

Correspondence to [Madina Mansurova](#).

## Editor information

---

Editors and Affiliations

**Yuri Gagarin State Technical University of Saratov, Saratov, Russia**

Olga Dolinina

**Saint Petersburg, Russia**

Igor Bessmertny

**Yuri Gagarin State Technical University of Saratov, Saratov, Russia**

Alexander Brovko

**Department of Computer Science, University of Texas at El Paso, El Paso, TX, USA**

Vladik Kreinovich

**Yuri Gagarin State Technical University of  
Saratov, Saratov, Russia**

Vitaly Pechenkin

**Yuri Gagarin State Technical University of  
Saratov, Saratov, Russia**

Alexey Lvov

**Technical Science, Novosibirsk Institute of the  
Software Systems, Novosibirsk Oblast, Russia**

Vadim Zhmud

Rights and permissions

---

[Reprints and Permissions](#)

Copyright information

---

© 2023 The Author(s), under exclusive license to  
Springer Nature Switzerland AG

About this paper

---

Cite this paper

Barakhnin, V., Mansurova, M., Grigorieva, I., Kozhemyakina, O., Ospan, A. (2023). TableProcessor: The Tool for the Analysis and the Interpretation of Web Tables to Create the Geo Knowledge Base of Kazakhstan. In: , *et al.* Artificial Intelligence in Models, Methods and Applications. AIES 2022. Studies in Systems, Decision and Control, vol 457. Springer, Cham. [https://doi.org/10.1007/978-3-031-22938-1\\_15](https://doi.org/10.1007/978-3-031-22938-1_15)

[.RIS](#)  [.ENW](#)  [.BIB](#) 

DOI

[https://doi.org/10.1007/978-3-031-22938-1\\_15](https://doi.org/10.1007/978-3-031-22938-1_15)

Published	Publisher Name	Print ISBN
25 April 2023	Springer, Cham	978-3-031- 22937-4

Online ISBN	eBook Packages
978-3-031- 22938-1	<a href="#">Intelligent Technologies and Robotics</a> <a href="#">Intelligent Technologies and Robotics (R0)</a>