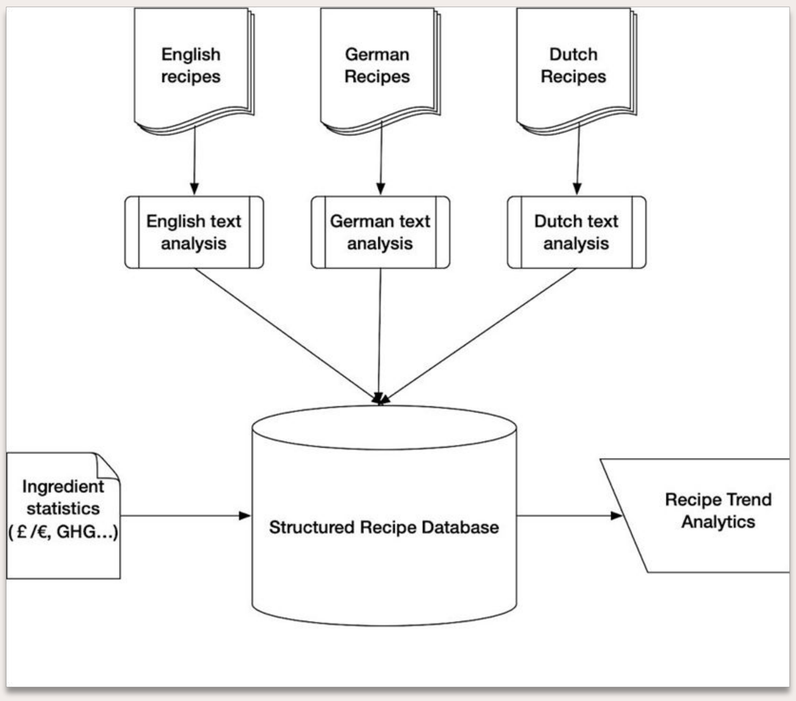
Comparing the environmental impacts of recipes from four different recipe databases using Natural Language Processing

Christian Reynolds, Berill Takacs, Anastasiia Klimashevskaia, Aslaug Angelsen, Rebeca Ibanez Martin, Steve Brewer, Marieke van Erp, Alain Starke, Diana Maynard, Christoph Trattner
christian.reynolds@city.ac.uk



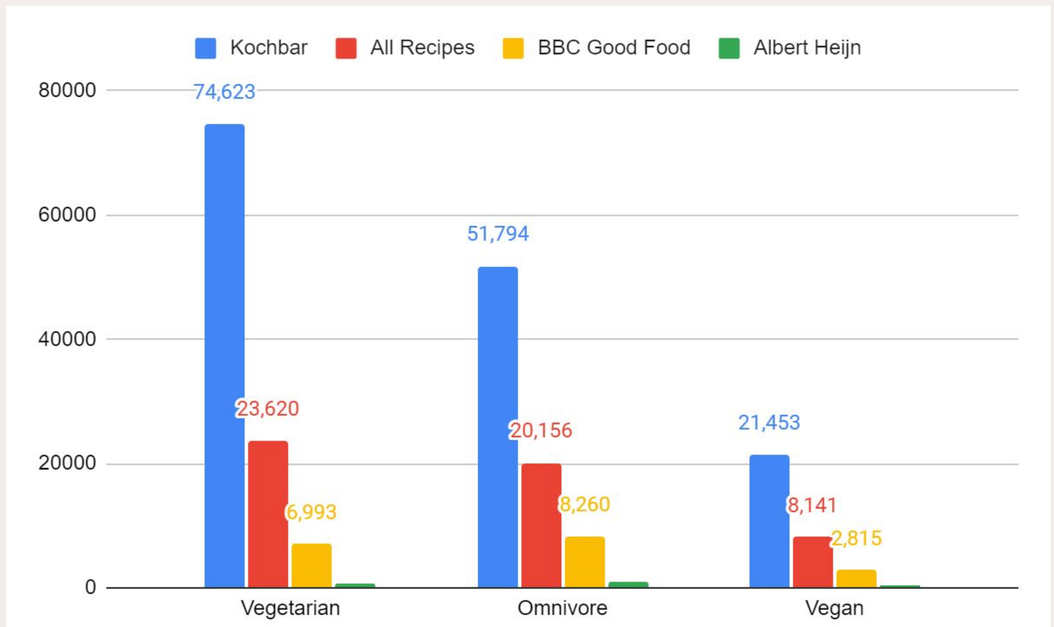


Figure 1. Number of Recipes per datasource, per diet

ABSTRACT

The calculation of environmental impacts from recipes remains a barrier to effective uptake of sustainable diets. In our project, we use pilot digital humanities methods to explore digitised recipe texts from websites in English, Dutch and German. Using the natural language processing toolkit GATE [1], we have developed customised tools to automatically extract ingredients, quantities and units from 220,168 Indexed recipes and match them to a food environmental impact database of 4500 ingredients (using the classification system FoodEx2). This database, based on environmental data from Poore and Nemecek (2018), provided Land Use (m2/FU), GHG Emissions (kg CO2eq/FU, IPCC 2013 incl. CC feedbacks), Eutrophying Emissions (g PO43-eq/FU, CML2 Baseline), Stress-Weighted Water Use (L/FU), and Freshwater Withdrawals (L/FU) for each ingredient. This allowed the calculation of these impacts at the mean, 5% and 95% confidence level per recipe and per portion. This has enabled us to explore the environmental impacts of vegan, vegetarian and non-vegetarian recipes if we were to cook these recipes using contemporary ingredients. To validate this tool we manually calculated the impacts of 50 recipes from 4 websites BBC Good Food, Albert Heijn/Allerhande, AllRecipes.com (Trattner et al 2017) and Kochbar (Trattner et al 2019) and compared these to the results from our tool.

[1] GATE is an open source software toolkit for automated text processing https://gate.ac.uk/overview.html

Nutrition information was sourced from the USDA FoodData Central (McKillop et al 2021) and McCance and Widdowson's Composition of Foods Integrated Dataset (Public Health England 2015). Environmental and Nutrition information was matched to two classification systems 4500 ingredients (FoodEx2 classification system) and 2842 ingredients (USDA Nutrient Database for Standard Reference, Release 24,classification system).

This poster fouses the differences in Median GHGE (Kg of Co2e) per Portion, based on the "diet" and the recipe datasource.

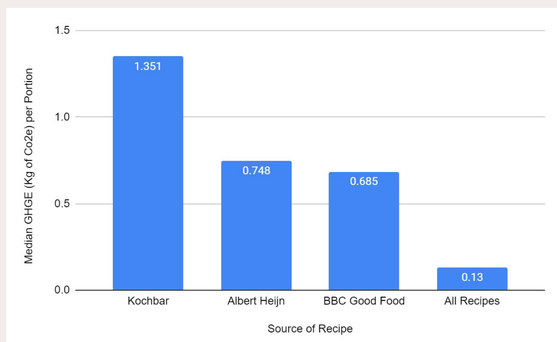


Figure 2. Median GHGE (Kg of Co2e) per Portion, per datasource

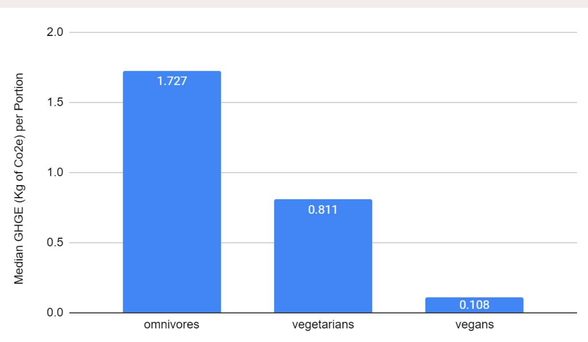


Figure 3. Median GHGE (Kg of Co2e) per Portion, per diet

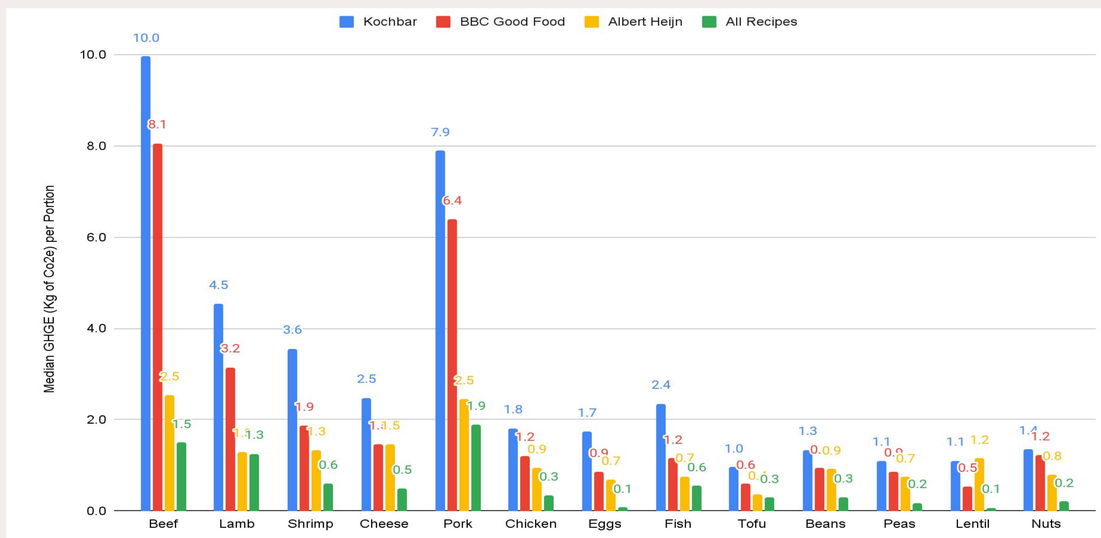


Figure 4. Median GHGE (Kg of Co2e) per Portion, Ingredient/Protein sources, and Datasource

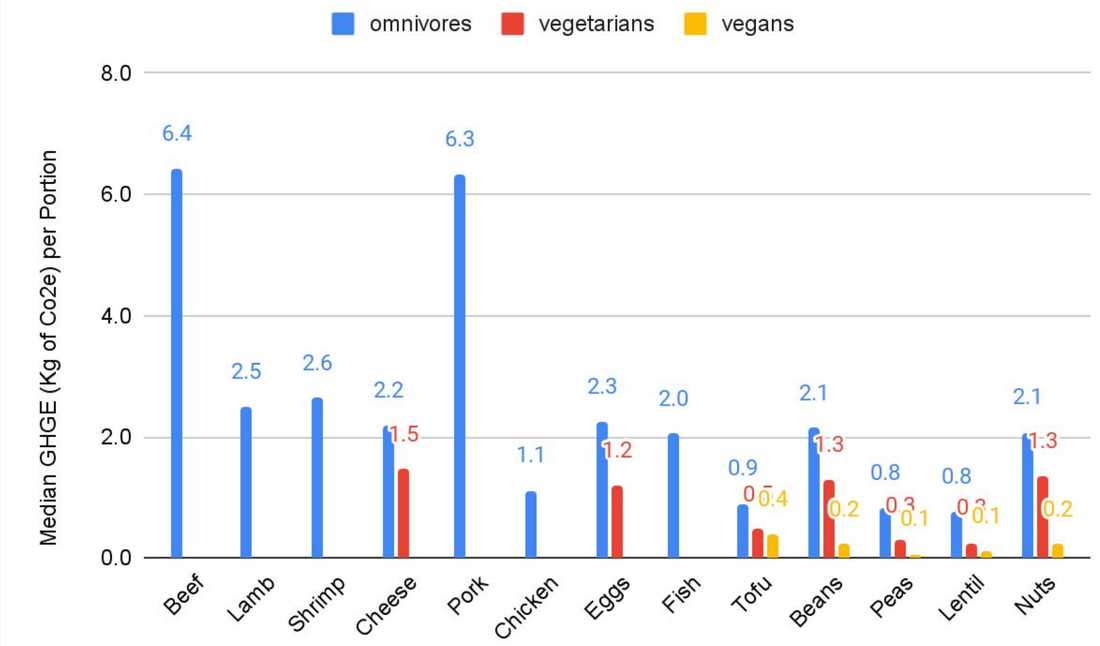


Figure 5. Median GHGE (Kg of Co2e) per Portion, Ingredient/Protein sources, and Diet

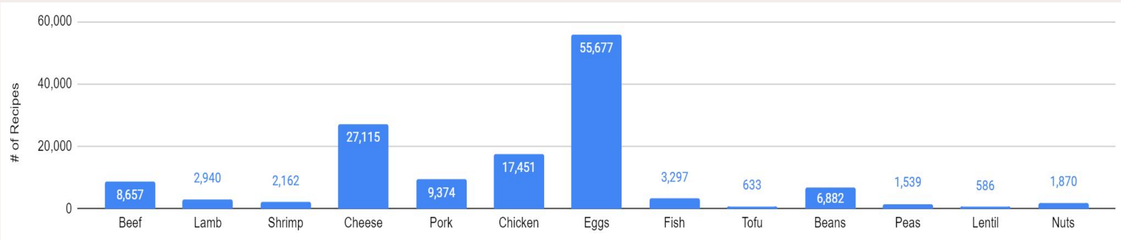


Figure 6. Number of recipes per Ingredient/Protein sources

REFERENCES

McKillop, K., Harnly, J., Pehrsson, P., Fukagawa, N. and Finley, J., 2021. FoodData Central, USDA's Updated Approach to Food Composition Data Systems. Current Developments in Nutrition, 5(Supplement_2), pp.596-596.
Poore, J. and Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. Science, 360(6392), pp.987-992.
Public Health England, 2015. McCance and Widdowson's composition of foods integrated dataset.
Trattner, C. and Elsweiler, D. (2017) Investigating the Healthiness of Internet-Sourced Recipes: Implications for Meal Planning and Recommender Systems. In Proceedings of the World Wide Web Conference (WWW).
Trattner, C., Kusmierczyk, T. and Norvag, K. (2019) Investigating and Predicting Online Food Recipe Upload Behavior. Information Processing and Management

ACKNOWLEDGEMENTS

This research and the GATE NLP tool have been developed with a research grant from the alpro foundation