

Ingredients matching in bakery products

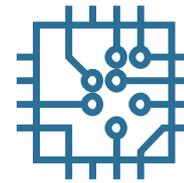
Tome Eftimov

Barbara Koroušič Seljak

<mailto:{tome.eftimov, barbara.koroušic}@ijs.si>



Safety
Quality
Traceability

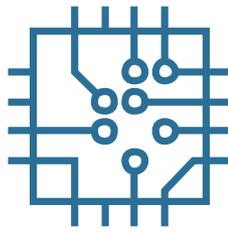


Computer
Systems

Jožef Stefan Institute

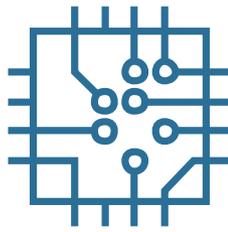


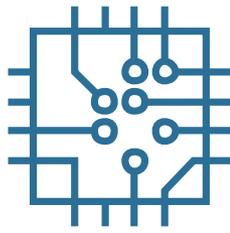
Objectives



- Relevant and interesting relations between recipes' ingredients
- Target data – bakery recipes

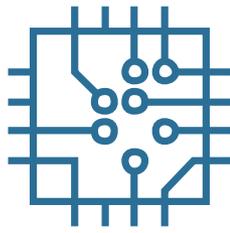
Finally settling down to my vegan, gluten free, soy free, antibiotics free, raw, non GMO, organic, fat free, low carb meal!



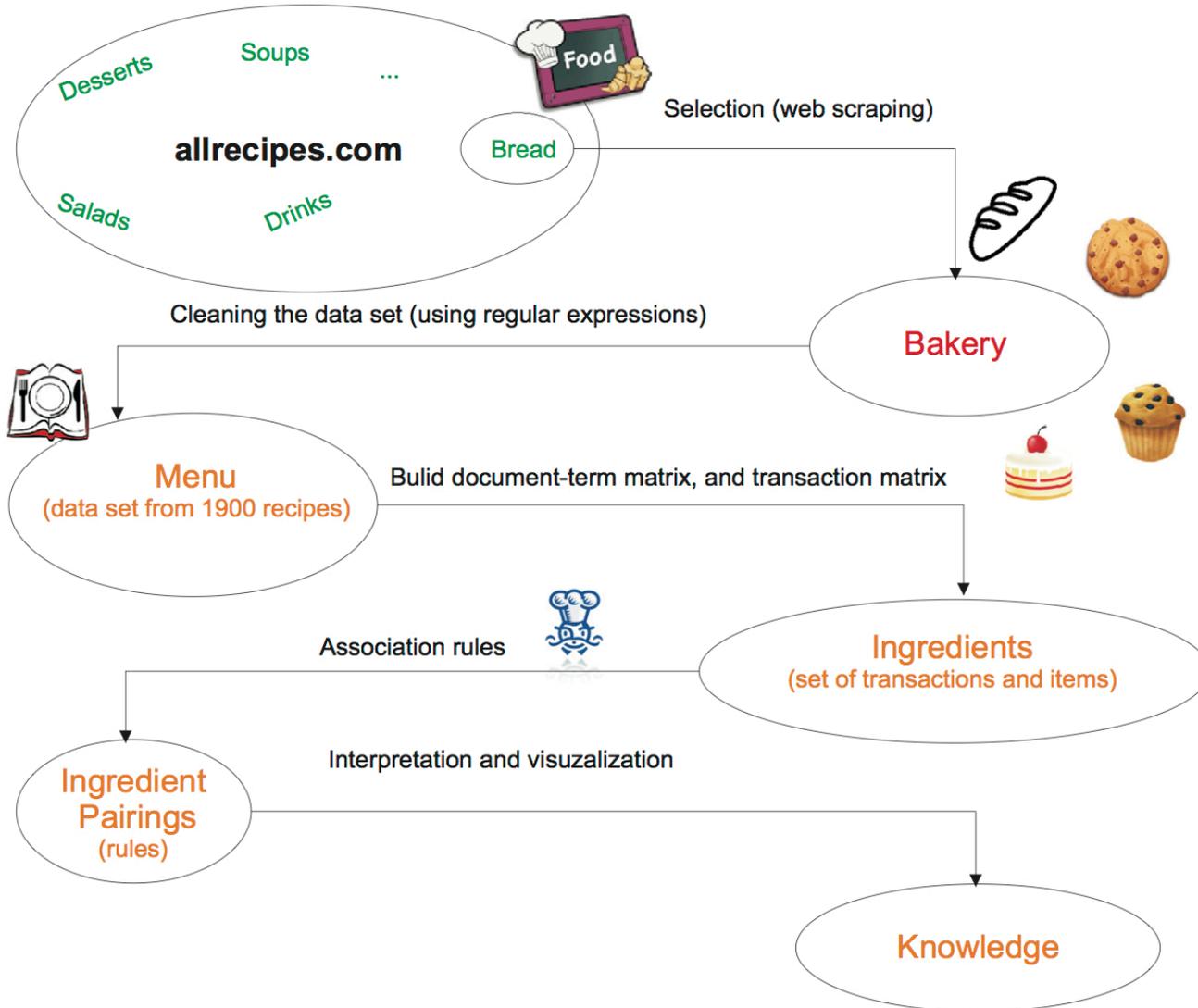


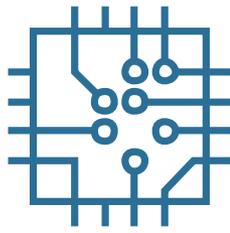
Introduction

- Data extraction
- Data preprocessing
- Association rules
- Visualization of the results



The knowledge discovery process

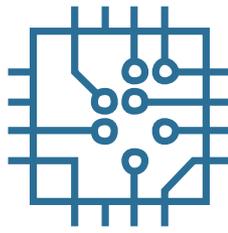




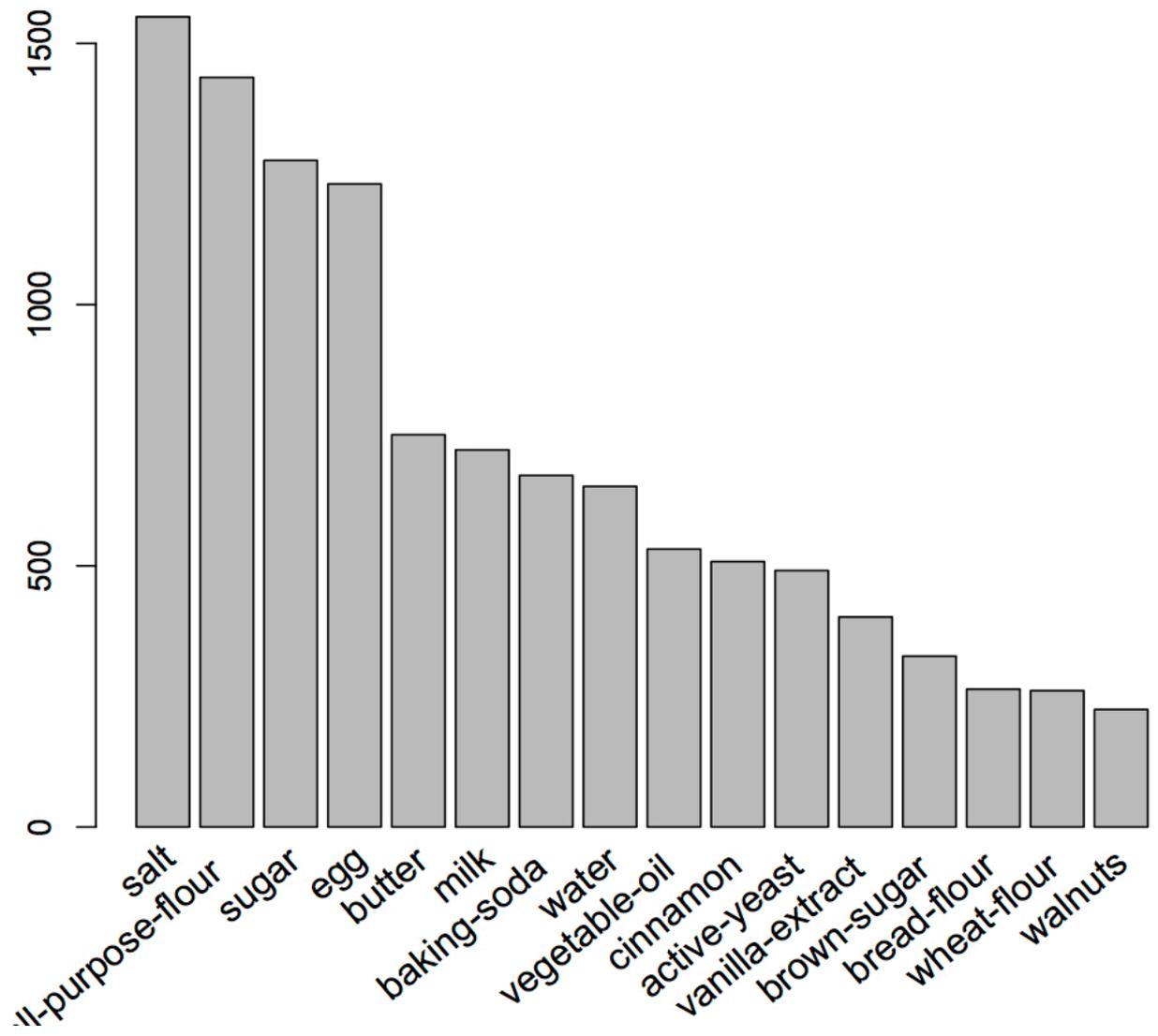
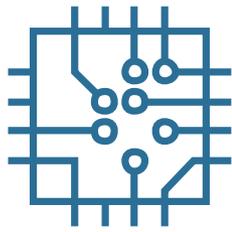
Data

- Collection of 1,900 bakery recipes written in English
- Data preprocessing: removed many adjectives associated with the cooking process, located synonyms
- Transformation: document-term matrix, transactional matrix

Basic statistics of the data set (1/2)

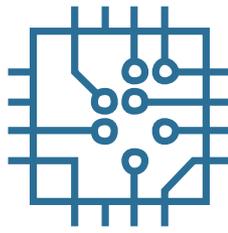


- Our transformed data contains 1,900 recipes and for each transaction we needed to consider the presence of 542 ingredients
- The data set is rather sparse with a density just above 1.65%
- The most popular ingredients are “salt”, “all-purpose flour”, “sugar” and “egg”

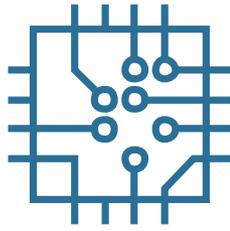


Basic statistics of the data set

(2/2)

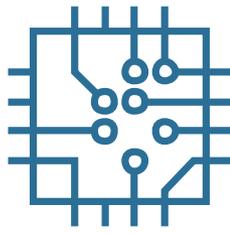


- We rejected the 4 most frequently used ingredients for the analysis
- Our data set contains 1,900 recipes and for each transaction we needed to consider the presence of 538 ingredients
- The data set is rather sparse with density just above 1.13%
- Average transaction contains less than 7 ingredients



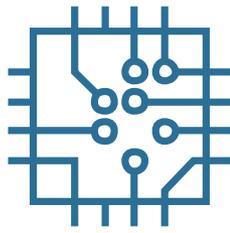
Methods

- Apriori algorithm
- Support, confidence, lift
- R programming language: “arules”,
“arulesViz”



Evaluation

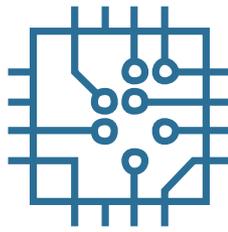
- Support = 0.005
- Confidence = 0.75
- The number of the discovered rules is 1,235
- After we pruned the redundant rules, the number of the discovered rules is 594



The top 15 discovered rules

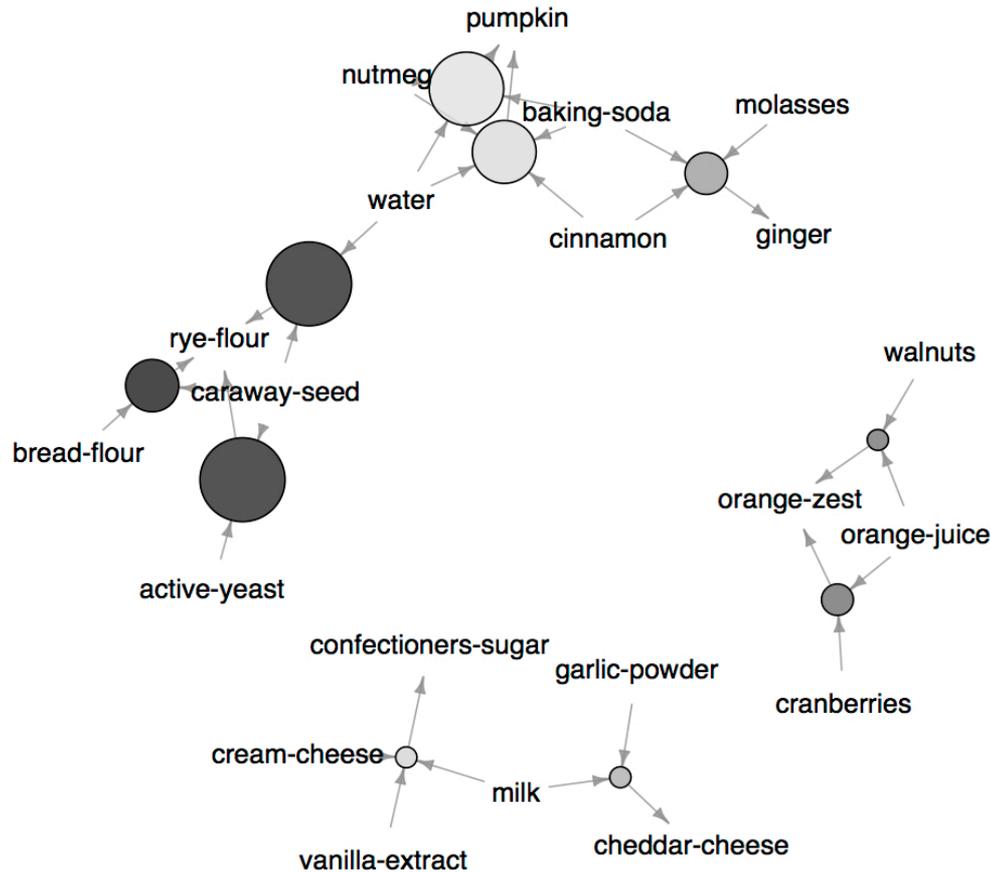
	LHS	RHS	support	confidence	lift
1	{bread-flour, caraway-seed}	{rye-flour}	0.006	0.928	45.238
2	{active-yeast, caraway-seed}	{rye-flour}	0.008	0.888	43.304
3	{caraway-seed, water}	{rye-flour}	0.008	0.888	43.304
4	{cranberries, orange-juice}	{orange-zest}	0.005	0.846	30.333
5	{orange-juice, walnuts}	{orange-zest}	0.005	0.833	29.874
6	{baking-soda, cinnamon, molasses}	{ginger}	0.006	0.800	24.126
7	{garlic-powder, milk}	{cheddar-cheese}	0.005	0.769	21.181
8	{cream-cheese, milk, vanilla-extract}	{confectioners-sugar}	0.005	0.909	16.142
9	{baking-soda, cinnamon, nutmeg, water}	{pumpkin}	0.007	0.823	14.901
10	{baking-soda, nutmeg, water}	{pumpkin}	0.007	0.789	14.285
11	{butter, cream-cheese, milk}	{confectioners-sugar}	0.005	0.785	13.951
12	{cinnamon, pumpkin-pie}	{pumpkin}	0.005	0.769	13.919
13	{allspice, water}	{pumpkin}	0.005	0.769	13.919
14	{pumpkin-pie, vegetable-oil}	{pumpkin}	0.006	0.764	13.837
15	{bread-flour, butter, water, wheat-flour}	{honey}	0.005	0.833	10.021

Graph visualization

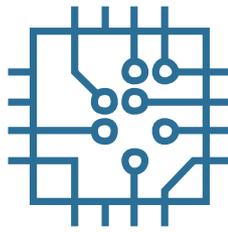


Graph for 10 rules

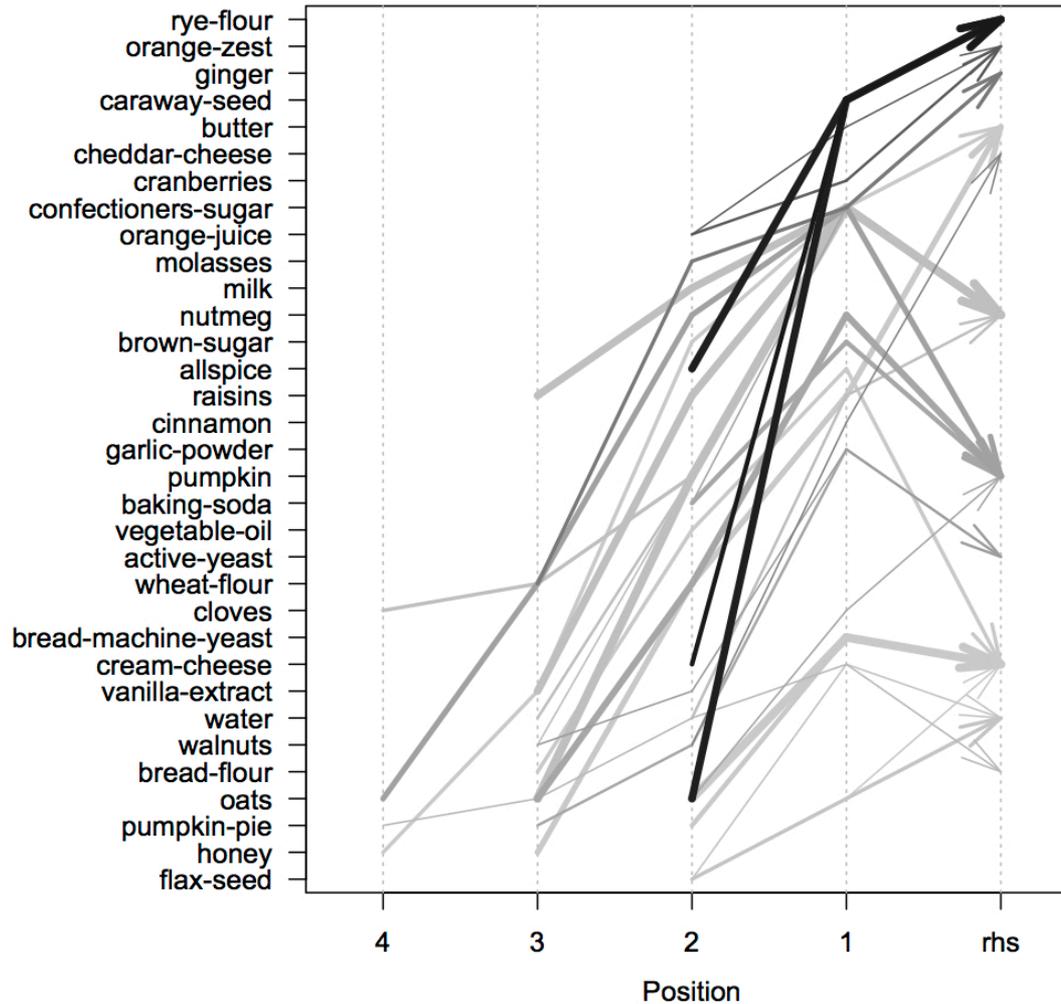
size: support (0.005 - 0.008)
color: lift (14.286 - 45.238)



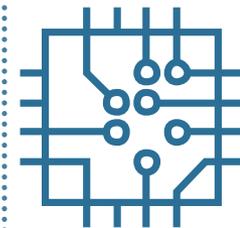
Parallel coordinates plot



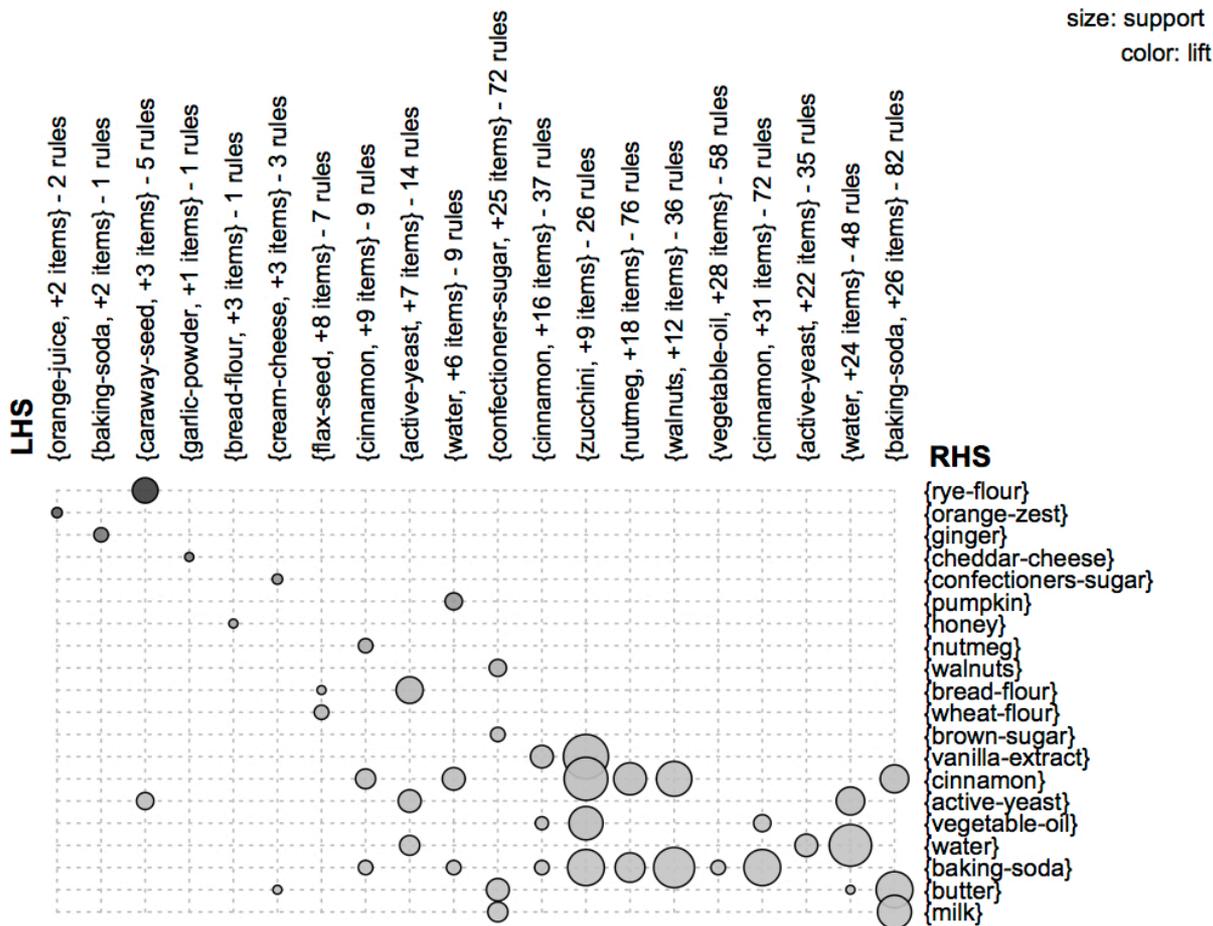
Parallel coordinates plot for 30 rules

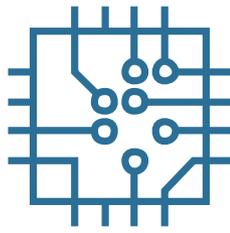


Grouped matrix visualization



Grouped matrix for 594 rules





Conclusion (1/3)

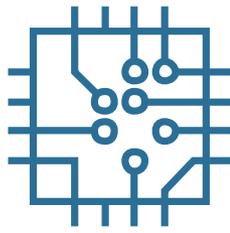
- Expected rules:

- {yeast} → {water}



- {apple} → {cinnamon}





Conclusion (2/3)

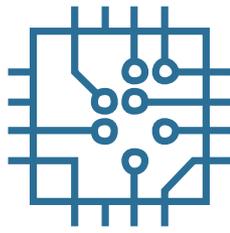
- Unexpected rules:

- {baking soda, cinnamon, molasses } → {ginger}



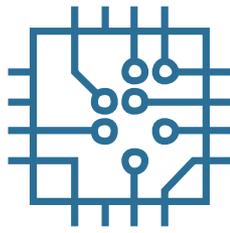
- {baking soda, nutmeg, water } → {pumpkin}





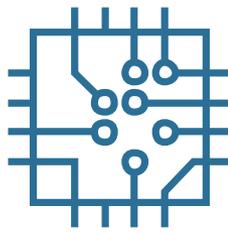
Conclusion (3/3)

- This analysis allows us to see how the ingredients are combined in bakery recipes
- The information is very important for food compilers who need to collect analytical data for food items frequently used in national dietary surveys based on foods and recipes



Further work

- To analyze these combinations in order to determine the nutritional properties for different values of quantity-unit pair for each ingredient
- Compare these relations with the relations provided by Foodparing



References

- Data Source:
http://allrecipes.com/Recipes/Bread/Main.aspx?prop24=hn_browsedeep&evt19=1
- Tan, Pang-Ning, and Vipin Kumar. "Chapter 6. Association Analysis: Basic Concepts and Algorithms." *Introduction to Data Mining*. Addison-Wesley. ISBN 321321367, 2005
- Rakesh Argawal and Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of 20th International Conference on Very Large Data Bases, vol. 1215, pp. 487-499, 1994
- Michael Hahsler, Christian Buchta, Bettina Gruen, Kurt Hornik and Christian Borgelt, "arules – Mining Association Rules and Frequent Itemsets", CRAN, 08.12.2014
- Michael Hahsler and Sudheer Chelluboina, "arulesViz – Visualizing Association Rules and Frequent Itemsets", CRAN, 08.12.2014