Shvetcova Karina, 014705566

Introduction to methods in digital humanities

Hands-on assignment

The aim of my research for master thesis is to study how the Russian cuisine is presented in the world by the example cookbooks abroad. In this pilot research I tried use digital humanities methods for processing of recipes and visualization of results.

My data

My pilot research based on cooking recipes of Borshch (Russian beet soup) in cookbooks in the English language. I chose this dish because it is symbolic dish of the Russian food culture. I found my data in OpenLibrary (https://openlibrary.org/). I chose several cookbooks about Russian cooking and borrowed them through Adobe Digital Editions:

"Allied Cookery British, French, Italian, Belgian, Russian" by Raoul Dandurand (1916)

"Cooking the Russian way" by Musia Soper (1961)

"The complete Russian cookbook" by Lynn Visson (1982)

"Please to the table" by Anya von Bremzen and John Welchman (1990)

"A taste of Russia" by Darra Goldstein (1999)

In all 23 recipes of Borshch from the cookbooks of different time periods.

Next I decided to preprocess my data and to do from them a structured data in order to simplify subsequent process. For that I did such table in Excel (**recipes_new.xlsx**)

The table includes columns with the name of cookbook, the author, the year of cookbook, conventional features from name of dish, name of dish and ingredients (where 1=yes, it includes, 2=no, no includes). In the end of ingredient columns I counted a frequency of each ingredient via the formula: = COUNTIF(F2:F24;"1")

And in the last column I counted number of ingredients in each recipe via the same formula.

Data processing

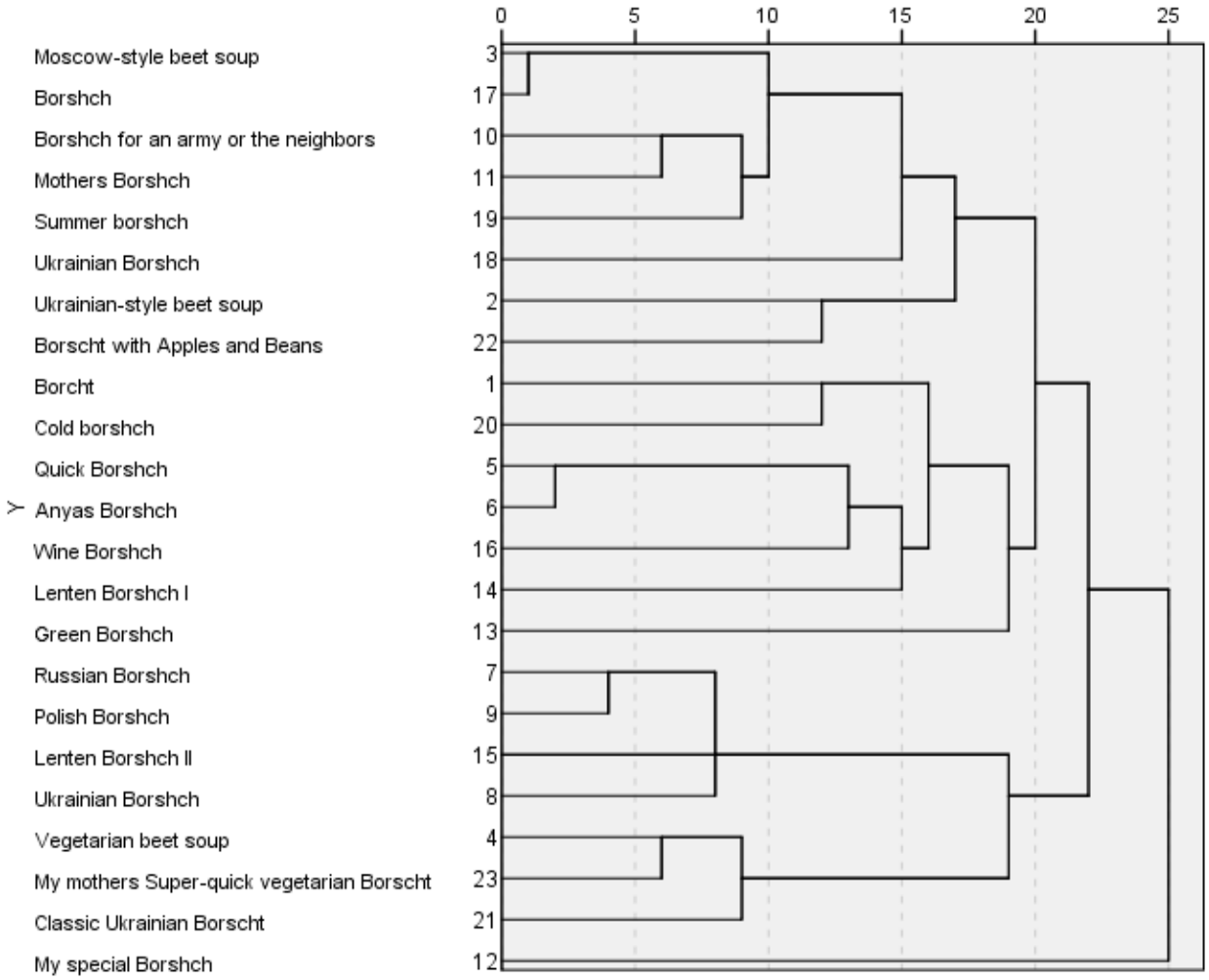
I wanted to study ingredients of recipes and their regularity therefore I decided to carry out a hierarchical clustering analysis with SPSS. For processing I transferred my data from Excel to SPSS (**ingredients.sav**).

I chose method of clustering analysis “Between-groups linkage”. After processing I received this table with division into clusters and dendrogram.

Cluster Membership

Case	7 clusters	6 clusters	5 clusters	4 clusters	3 clusters
1:Borcht	1	1	1	1	1
2:Ukrainian-style beet soup	2	2	2	2	1
3:Moscow-style beet soup	3	2	2	2	1
4:Vegetarian beet soup	4	3	3	3	2
5:Quick Borshch	1	1	1	1	1
6:Anyas Borshch	1	1	1	1	1
7:Russian Borshch	5	4	4	3	2
8:Ukrainian Borshch	5	4	4	3	2
9:Polish Borshch	5	4	4	3	2
10:Borshch for an army or the neighbors	3	2	2	2	1
11:Mothers Borshch	3	2	2	2	1
12:My special Borshch	6	5	5	4	3
13:Green Borshch	7	6	1	1	1
14:Lenten Borshch I	1	1	1	1	1
15:Lenten Borshch II	5	4	4	3	2
16:Wine Borshch	1	1	1	1	1
17:Borshch	3	2	2	2	1
18:Ukrainian Borshch	3	2	2	2	1
19:Summer borshch	3	2	2	2	1
20:Cold borshch	1	1	1	1	1
21:Classic Ukrainian Borscht	4	3	3	3	2
22:Borscht with Apples and Beans	2	2	2	2	1
23:My mothers Super-quick vegetarian Borscht	4	3	3	3	2

Dendrogram



Based on the table and the dendrogram I left only 5 clusters because in the case of 6 or 7 clusters the division has 2 clusters with one recipe. I suppose it is too detailed division.

Clusters:

1	2	3	4	5
- Borcht - Quick Borshch - Anyas Borshch - Green Borshch - Lenten Borshch I - Wine Borshch - Cold borshch	- Ukrainian-style beet soup - Borscht with Apples and Beans - Moscow-style beet soup - Borshch for an army or the neighbors - Mothers Borshch - Borshch - Ukrainian Borshch - Summer borshch	- Vegetarian beet soup - Classic Ukrainian Borscht - My mothers Super-quick vegetarian Borscht	- Russian Borshch - Ukrainian Borshch - Polish Borshch - Lenten Borshch II	- My special Borshch

Further I sorted out my data in Excel for analysis (**recipes_clusters.xlsx**).

My data are sorted out according to frequency of ingredient use in each cluster. As a result I received that:

- **First cluster** of recipes includes salt (100%), water (83%), onion (83%), carrot (83%), ground black pepper (83%), sour cream (83%), dill (67%), parsley (67%), beet (67%), potato (67%). Only 50% of all recipes in this cluster include a green cabbage.
- **Second cluster** of recipes includes salt (100%), ground black pepper (100%), beet (100%), carrot (88%), onion (88%), green cabbage (88%), bay leaves (88%), tomato (88%), water (75%), beef chuck (75%), celery (75%), sour cream (63%), potato (63%), vinegar (63%), bones (63%).
- **Third cluster** has in a composition: lemon juice (100%), bay leaves (100%), sour cream (100%), beet (100%), potato (100%), parsley (100%), dill (100%), green cabbage (100%), garlic (100%), sugar (100%), tomato paste (100%), black peppercorn (100%), salt (100%), ground black pepper (100%), carrot (100%), onion (100%), celery (67%), butter (67%), vegetable oil (67%), apple (67%), tomato (67%), green bell pepper (67%), prune (67%), water (67%).
- **Fourth cluster** has in a composition: salt (100%), carrot (100%), beet (100%), potato (100%), dill (100%), green cabbage (100%), lemon juice (100%), garlic (100%), leek (100%), parsnip (100%), water (75%), beef chuck or flank (75%), celery (75%), tomato (75%), tomato paste (75%). It doesn't include a sour cream in 75% of the recipes.
- **Fifth cluster** has only one recipe. It is recipe "My special Borshch" and it includes many special ingredients such as tomato sauce, eggs, string beans, peas, ketchup and brown sugar. It is difficult to assign it to one of other clusters. It is described in this recipe that "this very Russian borshch makes use of such American products as ketchup and brown sugar".

Analyzing

From a table with frequency (**recipes_new.xlsx**) we can see most common ingredients in Borshch: salt (in 22 from 23), beet (in 20 from 23), ground black pepper (in 20), carrot (in 19), onion (in 18), green cabbage (in 18), water (in 17), potato (in 16), sour cream (14), beef chuck, shank or flank (13), tomato (13). It forms the basis of soup Borshch.

Reasoning from the cluster division I gave a conventional name to each cluster: for first "Classic" (it includes common ingredients), for second "With bone broth and tomato but without fresh herbs", for third "Piquant" (many spices and aroma ingredients), for fourth "Without sour cream", for fifth "No cluster".

Also it is interesting fact that recipes with regional component in the name are not connected with ingredients. For example, "Ukrainian-style beet soup" from second cluster and "Classic Ukrainian

Borscht" from third cluster and "Ukrainian Borshch" from fourth. These clusters have their peculiarities and it is really strange fact that Ukrainian Borshch is in different groups.

After I added new column with the name of cluster in SPSS-table (**ingredients.sav**) for further research of Russian dishes in my master thesis.

Visualization

For visualization of my results I used Palladio. I changed values "1" and "2" in my Excel table to "Yes" and "No" respectively and added new column "Cluster". And as a result I received a visualization of clusters with name of recipes and dimensions according to ingredients number in a recipe. Also I can choose any ingredient and see where it is included and its value in percent.

Summary

The Borshch is in many cultures, for example in Polish, Ukrainian and Russian. This dish is integral part of Russian cooking and in concept of Russian person this soup should contain a beet and serve with sour cream. Based on this pilot research I can say that Borshch in cookbooks abroad has differences. It can be divided to clusters depending on ingredients and spices. But the name doesn't say us about this type of Borshch. Also using frequency analysis I determined the basis of Borshch:

- salt,
- beet,
- ground black pepper,
- carrot, onion,
- green cabbage,
- water,
- potato,
- sour cream,
- beef chuck, shank or flank,
- tomato.

These ingredients are almost in all recipes of Borshch.

In further research I would like to use this method and structure of study for other recipes of Russian cooking and to examine my theme comprehensively.