A Review of the Machine Learning Literature Relating to Digital Cooking Recipe Recommendation Service

Hasan Batuhan HASANOGLU, Didem ERDIVAN, Evrim Tugay MUTLU, Tolga TOLLUOGLU

c1511030@student.cankaya.edu.tr , c1511021@student.cankaya.edu.tr ,
c1411039@student.cankaya.edu.tr , c1511061@student.cankaya.edu.tr

Department of Computer Engineering, University of Cankaya

October 17, 2019, version 0.1

Abstract

Nowadays, most people are in a hustle. They have a busy work tempo or other daily hassles. So they neglect themselves. They don't even think about their health when preparing or eating food. At the same time, with developing technology, people become lazy. Even when making plans, they use these high-tech machines. For example in the evening, they even get help with what to cook at home. There are applications suggesting recipes for this. According to research on people, nowadays, time is more important than their health. Therefore, they prefer fast prepared meals like fast food. And this threatens their health. Or let's say we cook at home. We can easily reach the recipes we want. The problem is that we don't know what should we eat for our health and body. At the same time, this recipe must comply with our taste. At this point, our application is created as a solution to these problems. CookHub is a recipe recommendation Android application. You can find the recipes you are looking for. However, the app can also give you suggestions. But it doesn't make the same suggestions to everyone. Instead, we offer personalized meals that meet certain conditions. For example, the health status of the person, taste, already existing materials in the house. While we were preparing this report, we reviewed the articles about Machine Learning, AI and Data Science. We also searched articles on health and diet to classify recipes that fit the health status of users.

Keywords: Health, Recipe Recommendation, Machine Learning, Data Science, Android Application, Big Data analysis, Matching Algorithms

1. Introduction

With developing technology, people meet all their needs with technology. The most practical for this is the use of mobile applications. Nowadays, people care about their time and want to speed up their work. Therefore, they prefer to look at a recipe from applications rather than from a book. The data sets of the applications contain an incredible amount of data (millions of recipes). Users cannot decide which of these data to choose. CookHub is designed to help these people. We ask questions to these users at the registration step. We learn their health conditions, body mass measurements, the flavors they like to eat (roughly sweet, bitter, salty, spicy, etc.) via these questions. We prepare our recommendations based on this initial information. Later, users can evaluate recipes. We now consider these evaluations as we prepare new recommendations. Of course, we teach the algorithm to do this to our machine (application).

To provide diversity in our application, we have presented thousands of recipes to the user. We have done long dataset research. A dataset is a collection of data. In the case of tabular data, a dataset corresponds to one or more database tables. Each column has a certain variable and each row has a specific record. [27] We first tried to answer that question: "Where can we find datasets?" Then we came across a lot of websites. Like "FiveThirtyEight, BuzzFeed News, Kaggle, Socrata, Awesome-Public-Datasets on GitHub, Google Public Datasets, UCI Machine Learning Repository, Data.gov".

We chose to use "Kaggle". Kaggle enables users to find and publish data sets, explore and construct data science models in a web-based data science environment, work with other data scientists and machine learning engineers and enter competitions to solve data science challenges. [28] Most of the datasets are available to use. But it depends on the dataset. Some of the bigger companies have extra conditions that their data cannot be used without extra written permission.

The data set we use contains the following information; a step by step description of how to prepare and cook the food, ingredients in the food, calorie information, nutritional values, cooking time and so on.

We used MySQL to hold this large data set. We chose Python to use Machine Learning effectively. Besides, we will use Pandas, Mathplotlib, Numpy, which are important Python libraries for us. Pandas is a software library for the Python programming language. it is used in computer programming, data processing, and analysis. First, it provides data structures and operations for modifying tables and time series. [1] Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. We can generate plots, histograms, power spectra, bar charts, error charts, scatterplots, etc [2] NumPy (Numerical Python) is a mathematical library that allows us to perform scientific calculations quickly. It is based on numpy arrays. [3]

We will use these libraries together. So we can process our large data set. When we recommend new recipes to users, we consider based on their previously preferred recipes and specified health status. We describe the characteristics of the recipes with short tags. For example, spicy, with eggs, lactose-free… We also keep these tags in the data set. These tags are made from the best match to users. We scale the matching rate of the newly proposed recipe and user information with a term called "accuracy".

The most distinctive point of our application is to care about the health of the user while finding personalized recipes. For example, a user with a lactose allergy adds this to his / her information when registering. So we offer her recipes with the label "lactose-free" while we suggest a recipe. We ask a lot of questions when registering for users who want to do a diet. So we want to provide him with the most suitable diet. "Who should consume what? Who should avoid what? Who should take how many calories per day?" We aim to be a recipe app that can help you to answer questions such as.

Another feature of the application is a screen where you can click on the materials already available in your home. These materials of your choice will be matched to the tags in the data set in the background. And you will be offered what you can do with the materials in your home.

We also consider users' health conditions. We could not do this without using certain mathematical formulas and research on this topic. To get the best efficiency in calorie and diet, we examined similar applications that serve the same purpose. We found that all of them used a BMI Calculator. So we will use it.

Height and weight are an important variable in nutrition. Metrics that adjust the weight for height, weight, and weight-related health risks and mortality. Ideal body weight (IBW) 2 was considered a "healthy" weight. The right conditions for BMI are:

1) Adult weight increases proportionally to height squared

2) BMI is a stature-independent measure of weight

3) BMI has a good correlation with fat mass (i.e., population studies the r values are ~0.7) [4]

2. Related Works

A recommender feature refers to a device that can forecast a user's potential choice for a collection of products, and suggest the top items. A key reason why we need a recommender program in modern society is because users have too many choices to use because of the Internet's proliferation. So we will use the three machine learning algorithms about the recommender system. These algorithms are Singular Value decomposition (SVD) - Jaccard Similarity Coefficient - Naive Bayes.

The Jaccard Similarity Index (sometimes referred to as the Jaccard Similarity Coefficient) compares members for two sets to see which members are shared and which are separate. It is a test of correlation for the two data sets, varying from 0 percent to 100 percent. The higher the percentage, the more similar the two populations are. While simple to interpret, it is extremely sensitive to small sample sizes and may provide incorrect results, particularly with very small samples or data sets with missing observations.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Figure (a): Jaccard similarity is computed using the following formula

The Naive Bayes algorithm is a basic probabilistic classifier, calculating by counting a collection of probabilities a given data set's frequency and value combinations. The algorithm uses Bayes theorem which claims that, given the value of the category function, all attributes are equal. This contingent presumption of freedom seldom occurs to real-world applications, hence the description as Naive yet the algorithm continues to do well and learn quickly in various supervised classification problems. Classification algorithm performance is usually examined by evaluating the classification accuracy. Since classification is often a confusing issue, the correct answer can depend on the user. Standard analysis methods to algorithms such as evaluating the

overhead space and time can be used, but these strategies are typically secondary. Evaluating which is better relies on consumers understanding the problem.

SVD is used as collaborative filtering (CF) algorithm in the context of recommendation systems. Collaborative filtering is a method for predicting a user-item pair rating based on the user's history of ratings given to the item. Most CF algorithms are based on a user-item rating Each column represents an item in a matrix where each row represents a user. The entries in this matrix are ratings that users give to items.

Collaborative filtering(CF) is a technique that can be used by recommendation systems. Collaborative filtering in the newer, narrower sense is a form of automated predictions(filtering) concerning the user's desires by gathering tastes or taste data from many users(collaborating).

Cooking is an art and it has to be performed by rules to success. Also, another important issue is the recipe for the food. These are the rules of cooking a portion of food. But how about health and taste? Not all foods are for all people's taste and health. Why not combine machine learning and cooking for the aim of health and perfect palatal delight.

There is already a lot of services to search for recipes. For example, Google Recipe Search [5], Yummly, Foodily, and MyTaste… Good results can be achieved by using these services. Because these services have many search capabilities. This makes it easy to use. For example, applications have search limits such as ingredients, species, cooking time, portions and nutritional values. Some of this information is marked as machine-readable. [6]

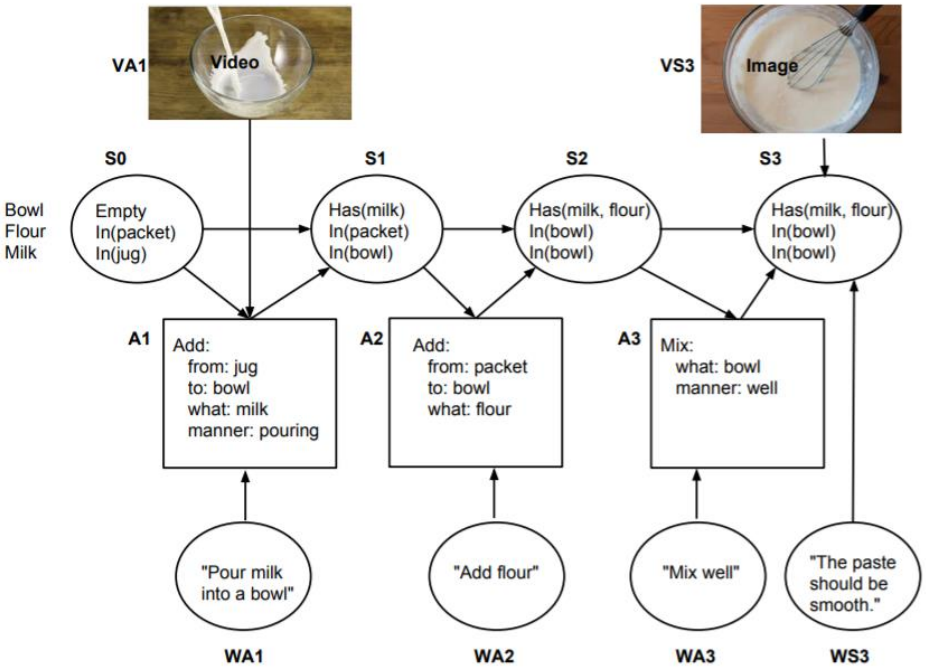

Figure (b): the actual steps of the recipe [7]

A machine can not recommend every food to every human group. Everybody has different concerns like health, gluten-free, calorie level, taste. For this reason, the machine has to learn these issues about each user group interactively. By this machine learning system recommendation system works more successful as the groups use the system. Taste, health concerns, etc. are detected by tags of the data in data. Furthermore, we help people to plan their meals without eating foods they don't like while they are improving their health and life standards.

Despite technological advancement, people are still not implementing healthy diet programs. It is possible to find diet programs from books, websites, and mobile applications. However, people prefer these programs to be personal. Finding such programs is not easy. So, the food mentioned in any diet program is not attractive. And users will not strictly follow the diet rules. To solve this problem, they create personalized meal plans for suggestion systems. Studies on this subject have focused on general users. One of the main considerations when designing a proposal is that food and consumption are recurring. People eat at similar times a day and every day, and they plan their meals sequentially. A smart machine can help to pick healthy and tasty foods for them. [8]

The usage of the internet makes people's life easy in anyways. Especially in needs like food. Recipe recognition with a large multimodal food dataset is very important because the main challenge is the size of the data set. Classification of food types is a crucial technology for many food related applications, such as healthy diet control, automated cooking, meal advisory program, etc. Throughout, a novel smartphone application is being built to monitor daily meal events using memory recovery technique. Based on this unique dietary data log file, they were able to carry out further studies on differential use and measure the balance of the food intake.

The Open Meal System aims to invent new intelligent cooking appliances. This is a Purdue University Technology-Assisted Diet Assessment (TADA) project. The project aims to develop a mobile food recorder. This device records diet information. The device plans this information daily. It considers daily nutritional intake. The food category classification should be in all these applications. [9]

The recommendation system is a relative concept. Netflix is a good example of a recommendation system. To date, Netflix has a bit over 100 million users on its site. But with every account having multiple accounts, this is double the actual number. That's a lot of data, and with more than 80 percent of TV shows on the platform being viewed, it's got its work cut out to keep people interested.. But not all of the hard work has to be finished.

Effectively, Netflix outsources certain processes to you. Netflix collects and analyzes information on how you use the platform, how you rate content, and what you're looking for to improve your user experience.

When you access the Netflix service, a suggestion system is activated. The main purpose of the system is to show you the ideal show or film. We can think of Netflix as a film catalog. In this catalog, movies that you are likely to watch are presented according to the following factors:

- Your interactions with the Netflix service (such as your viewing history)
- Other members with similar preferences
- Types, categories, players, year of publication and so on. Information about topics

Netflix knows what to watch with these factors. It also looks at the following to personalize a similar and best suggestion:

- The time of day you watch

- The devices you are watching Netflix on
- How long you watch.

All of these factors mentioned above are kept as data. These data are the inputs of the Netflix algorithm. (The algorithm is the sequence of steps and processes followed to solve a problem.) Netflix does not take age and gender into consideration when deciding on the most appropriate recommendations. [10]

The Effectiveness of a Smartphone Application is an important concept for diet and a healthy life. On 10 October 2018, José I Recio-Rodríguez and his team published an article about The effectiveness of a mobile program to alter primary care intakes of macro and micronutrients. The aim of this study is to evaluate the long-term (12 months) effectiveness of adding a diet smartphone application to standard dietary composition modification counseling (macro and micronutrients and food groups). According to this research consisting of nutritional counseling and a diet smartphone application, achieved better results than counseling alone in modifying the diet.[11]

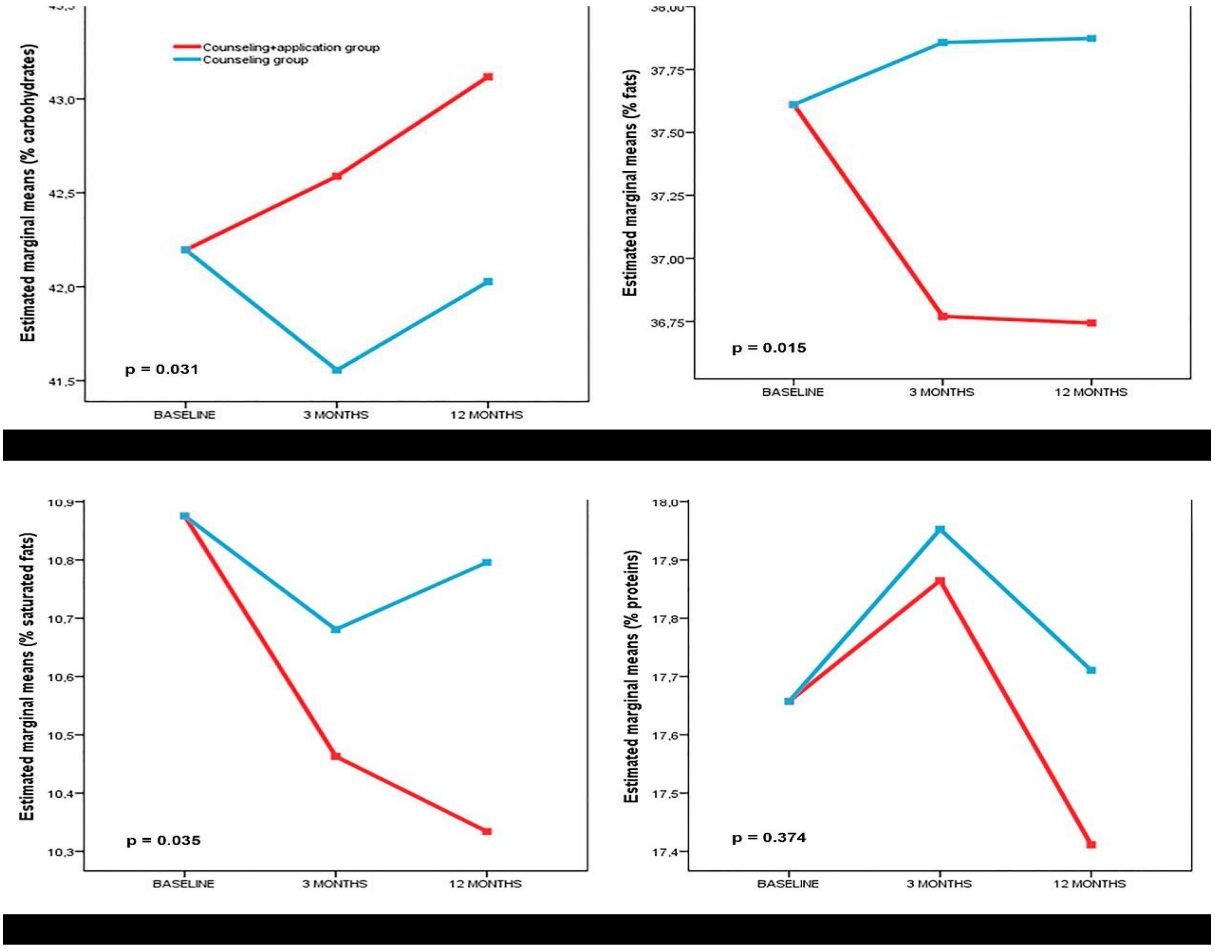

Figure (c): Changes in macronutrients after randomization by group (repeated measures analysis). [11]

Rapid technological developments have encouraged the use of smartphones in research and practice related to health promotion. Even though many diet- and nutrition-related applications (apps) are available from major smartphone platforms, relatively few research studies have been tested to determine their effectiveness in health promotion. In qualitative research, participants favoured fast and easy-to-administer applications and those that raised awareness of food intake and weight

management. In randomized trials, the use of smartphone apps was associated with better dietary compliance for lower-calorie, low fat, and high fiber foods, and higher physical activity levels which resulted in more weight loss.

Smartphone apps are expected to be a cost-effective and useful tool to boost diet and nutrition and tackle obesity in the general population. The precision of measures of the diet and nutrition achieved using mobile devices was generally found to be high. Participants prefer quick and easy-to-administer applications and those that increase awareness of food intake and weight management. [12]

The development toolkits and programming paradigms used by application developers have a significant impact on the architecture design and user interface of the apps that they are designing. Particularly now that technological advances provide us with inexpensive hardware capable of both sensing and viewing information with unparalleled precision, vividness, and bandwidth, there is an increasing need to explore and develop tools that allow innovative designers and developers to make their ideas a reality. These tools must both enable rapid prototyping to quickly evaluate new ideas and be able to leverage the full computing power of cutting edge technology. Python Kivy framework is one of them. [13]

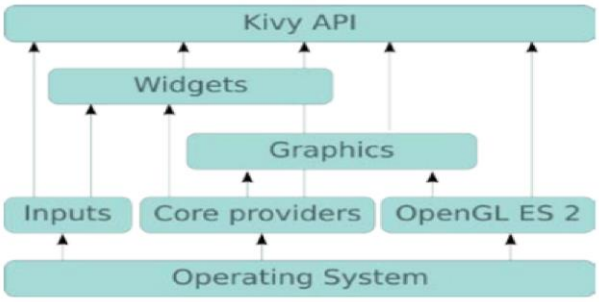

Figure (d): A general architecture of Kivy.[13]

A culinary life is now attracting attention as a consequence of the lifestyle-related disease crisis. Good eating habits are essential to a healthy life. Nonetheless, menu preparation involves consideration of various factors, such as nutritional value, in-store items, food preferences and expense. People therefore need to expand considerable effort to plan their daily menu. Against this backdrop, numerous cooking websites have recently been introduced, offering various food recipes such as Cookpad[14.1] and AllRecipes[14.2]. When preparing their meal several people refer to these websites. Cookpad has 900,000 recipes, and 10,000,000 users per month. This data reflects the high demand for services to provide the recipes. These websites do not, however, reflect the preferences and conditions of the user, although these two factors must be considered if the aim is to provide highly satisfactory recipes.

In January 2011, Mayumi Ueda and her team published an article about this subject. They presented a method in their paper to extract the food preferences of the user for recommendation of a recipe. Their approach predicts the interests of a person through his / her past actions, for example from their experience of searching the recipe and menu planning. To extract the preferences, their method breaks down the recipes into their ingredients and scores the recipes using ingredient frequency and specificity. As their method can estimate preferences through their history of browsing and cooking, the user transmits his / her preferences to the system without having to perform any particular operation. [14]

Another application that has parallel functions with us is Kochbot. Kochbot is a cooking assistant application for smartphones and tablet devices. This application allows you to search from a wide collection of recipes and explains cooking commands step by step. The most important feature of this

application is that it is focused on speaking and listening. The main advantage of the application is that it has a hands-free scenario. and user satisfaction is quite high. [15]

The Internet provides us many knowledge-sharing facilities and searching for any information from any corner of the world. Recipe websites are good examples of this knowledge sharing. People can find millions of recipes using the internet. The many recipes that the internet provides are kinda difficult for a person who just want to search a recipe fits in his taste. On the Internet, most of these websites are simple-minded as not reliable on recommendation systems that recommended recipes based on rating and comment-based recipes on the site. In these cases, a person who has been looking for a recipe for his taste can't always happy about what they get.

Bon Vivant is an interactive platform for discovering flavor mappings based on flavor compound analysis. It makes combinations of meals. It can match many factors, being able to do so. Eg ingredients, recipes, regional kitchens, and aroma compounds. This platform enables the user to achieve a healthier lifestyle. Because it is based on nutritional requirements when making recommendations. [16]

Smart mobile phones, which have entered almost every home in our age, do many things to make our lives easier. Good eating habits are key for a healthy life. How ever food planning requires consideration of different factors, which are nutritional value, in-store dietary, food preferences and expense Additionally, people need to expand a lot of effort and time towards planning their daily food cooking. For these subjects, various websites provide various food recipes that have been published worldwide such as Cookpad[14.1] and Yahoo! Recipe[17]. Many people refer to these websites to answer their "how to cook" question. Cookpad contains over 100 million people worldwide now use Cookpad every month, and more than 4 million recipes were produced on the website. It’s available in almost 70 countries around the world in 23 different languages. [18]

3. Conclusion

People are very busy in their daily lives. They want to get things done quickly. They even want speed while thinking. And, of course, when deciding. Sometimes they can't decide what to eat. They don't even know what should/shouldn’t eat. That’s why we decided to create an application that recommends both healthy and personalized recipes and shows you how to make it step by step. We have designed many algorithms to create this application. We have examined similar works from various articles. And we mentioned them in our report.

4. References

[1] https://pandas.pydata.org/

[2] https://matplotlib.org/

[3] https://numpy.org/

[4] *The American Journal of Clinical Nutrition*, Volume 103, Issue 5, May 2016, Pages 1193–1194,

Available: https://doi.org/10.3945/ajcn.116.134221

[5] http: //www.google.com/insidesearch/ features / recipes /

[6] http://microformats.org/wiki/ recipe formats

[7] https://www.aclweb.org/anthology/W14-2407/

[8] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.666.1434&rep=rep1&type=pdf

[9] https://ieeexplore.ieee.org/abstract/document/7169757

[10] https://help.netflix.com/en/node/100639

[11] The Effectiveness of a Smartphone Application on Modifying the Intakes of Macro and Micronutrients in Primary Care: A Randomized Controlled Trial. The EVIDENT II Study (10 October 2018)

[12] Steven S Coughlin, Mary S. Whitehead, Joyce Q Sheats, January 2016, Smartphone Applications for Promoting Healthy Diet and Nutrition: A Literature Review

[13] Kivy – A Framework for Rapid Creation of Innovative User Interfaces

Available: https://www.semanticscholar.org/paper/Kivy-A-Framework-for-Rapid-Creation-of-Innovative-Virbel-Hansen/bae0df60195a86fa7fd12172b026b32ff7cfbe1b

[14.1] Cookpad. http://cookpad.com/ , (Accessed 20 October 2019)

[14.2] AllRecipes. https://www.allrecipes.com/ , (Accessed 20 October 2019)

[14] M. Ueda, M. Takahata, and S. Nakajima. User's food preference extraction for cooking recipe recommendation. In Proc. of the 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation, 2011

[15] Ulrich Sch¨afer, Frederik Arnold, Simon Ostermann, and Saskia Reifers, 2013, Ingredients and Recipe for a Robust Mobile Speech-Enabled Cooking Assistant for German

[16] http://www.oaijse.com/VolumeArticles/FullTextPDF/224_40.BON_VIVANT_AN_ARTIFICIAL_INTELLIGENCE_COOKING_APP.pdf

[17] https://loco.yahoo.co.jp/gourmet/recipes/

[18] https://medium.com/cookpadteam/cookpad-the-story-behind-the-platform-used-by-100-million-people-7060f7fa4833

[19] Brants, T.: TnT – A statistical part-of-speech tagger. In: Proc. of 6th ANLP, Seattle, Washington, pp. 224–231 (2000)Google Scholar Hamada, R., Okabe, J., Ide, I.: Cooking navi: Assistant for daily cooking in kitchen. In: Proc. of 13th ACM Int. Conf. on Multimedia, Singapore, pp. 371–374 (2005)Google Scholar

[20] Chouambe, L.C.: Dynamische Vokabularerweiterung für ein grammatikbasiertes Dialogsystem durch Online-Ressourcen, Studienarbeit, University of Karlsruhe (2006)Google Scholar

[21] Drozdzynski, W., Krieger, H.U., Piskorski, J., Schäfer, U., Xu, F.: Shallow processing with unification and typed feature structures — Foundations and applications. Künstliche Intelligenz 1, 17–23 (2004)Google Scholar

[22] Hamada, R., Okabe, J., Ide, I.: Cooking navi: Assistant for daily cooking in kitchen. In: Proc. of 13th ACM Int. Conf. on Multimedia, Singapore, pp. 371–374 (2005)Google Scholar

[23] Martins, F.M., Pardal, J.P., Franqueira, L., Arez, P., Mamede, N.J.: Starting to cook a tutoring dialogue system. In: SLT Workshop 2008, pp. 145–148. IEEE (2008)Google Scholar

[24] Petitpierre, D., Russell, G.: MMORPH – the Multext morphology program. Tech. rep., ISSCO, University of Geneva (1995)Google Scholar

[25] Ribeiro, R., Batista, F., Pardal, J.P., Mamede, N.J., Pinto, H.S.: Cooking an ontology. In: Euzenat, J., Domingue, J. (eds.) AIMSA 2006. LNCS (LNAI), vol. 4183, pp. 213–221. Springer, Heidelberg (2006)CrossRefGoogle Scholar

[26] Wasinger, R.: Dialog-based user interfaces featuring a home cooking assistant, University of Sydney, Australia (2001) (unpublished manuscript)Google Scholar

[27] https://en.wikipedia.org › wiki › Data_set

[28] https://en.m.wikipedia.org › wiki › Kaggle