

CookIT: A Web Portal for the Preservation and Dissemination of Traditional Italian Recipes

M. T. Artese, G. Ciocca, I. Gagliardi

Abstract—Food is a social and cultural aspect of every individual. Food products, processing, and traditions have been identified as cultural objects carrying history and identity of social groups. Traditional recipes are passed down from one generation to the other, often to strengthen the link with the territory. The paper presents CookIT, a web portal developed to collect Italian traditional recipes related to regional cuisine, with the purpose to disseminate the knowledge of typical Italian recipes and the Mediterranean diet which is a significant part of Italian cuisine. The system designed is completed with multimodal means of browsing and data retrieval. Stored recipes can be retrieved integrating and combining a number of different methods and keys, while the results are displayed using classical styles, such as list and mosaic, and also using maps and graphs, with which users can play using available keys for interaction.

Keywords—Collaborative portal, Italian cuisine, intangible cultural heritage, traditional recipes, searching and browsing.

I. INTRODUCTION

FOOD is an important part of everyone life. Every culture has its own products, preparations and traditions. Traditional cuisine is passed down from one generation to the next and, often, it is an expression of cultural identity. For example, migrants bring the food of their countries with them and cooking traditional food is a way of preserving their own culture outside their birth places. It is not surprising that nowadays, food is receiving particular attention from government agencies with the aims of preserving historical roots and cultural identification, sourcing food as near as possible to the production farms, and coping with health issues such as reducing obesity and unbalanced nutrition. Also, UNESCO included food in its Intangible Cultural Heritage list, such as "gastronomic meal of the French", the "traditional Mexican cuisine", the "Nsima, culinary tradition of Malawi", and the "Mediterranean diet".

The system CookIT is presented here, a web portal related to Italian Traditional Recipes, with the aim of preserving and disseminating them. Furthermore, the dataset is developed with the aim to be used to test new algorithms for content-based retrieval on text and images. The collection is still under construction.

The paper is structured as follows: after a brief overview of

the most important cuisine websites, the main features of the portal are listed, together with the design methodology used, the creation of the dataset, the innovative navigation and retrieval methods implemented. Finally, future developments and conclusions are presented.

II. RECIPES PORTALS

Searching on the web terms like 'recipes' or 'step by step recipes' or 'traditional recipes' thousands of sites can be found. The content and purpose of these sites can be of various kinds. The purpose may be, for example:

- to share traditional or innovative recipes, sometimes giving image step-by-step directions, such as [1], [2];
- to give a taste of the recipes contained in the magazines or shown on TV as 150 Classic recipes of Savour [2], British recipes of BBC GoodFood [3], Telegraph [4], or Cooking channel TV [5];
- to advertise a brand, a store or a chef [6]-[8];
- to spread special diet recipes as healthy [11], gluten free [9], vegan [10] or dairy-free;
- to create a community on food and recipes [12]-[14].

Wikimedia project Wikibook has a section related to Cookbook, with the aim of sharing "A growing, world-wide, collection of recipes, translated into, or written in, English with links to other languages", as stated in the tagline [15].

A great number of Italian recipe sites or blogs are available on the web, in English or Italian language. The websites can be divided into two broad categories, as regards the set of recipes contained: 1) large generalist sites that contain both traditional recipes and innovative recipes, along with tutorials, tips, etc. [16]-[22]; 2) very local small sites, which describe culinary traditions such as Liguria or Abruzzo, two Italian regions [23]-[25].

Even though there are many websites dedicated to recipes, a new site and dataset has been created because: 1) recipes must be only those of the Italian culinary tradition, specifically related to all Italian regions, 2) it was necessary to have a supervised dataset, to allow the application of new visualization and search algorithms.

III. COOKIT PORTAL

CookIT is a Web portal dedicated to recipes of traditional Italian cuisine. Its goal is to spread the knowledge of Italian food recipes and the Mediterranean diet of which the Italian cuisine is an important part. However, Italian cuisine is variegated with thousands of different recipes that are ever changing. For this reason, the portal focuses only on traditional food recipes, considered the core and hearth of the

M.T. Artese and I. Gagliardi are with the Institute for Applied Mathematics and Information Technologies "Enrico Magenes", National Research Council of Italy, Milano, MI 20133 Italy (e-mail: teresa@mi.imati.cnr.it, isabella@mi.imati.cnr.it).

G. Ciocca is with the Department of Informatics, Systems and Communications, University of Milano Bicocca, Milano, MI 20133 Italy (e-mail: ciocca@disco.unimib.it).

Italian cuisine.

The portal has been designed with the aim of creating a site that is efficient, effective, usable and user-experience oriented [26], [27]. Main characteristics of the portal:

- 1) Multimedia Information system oriented towards traditional Italian recipes, in Italian language; targeted to people interested to discover real traditions about food and procedures to cook it, together with historical background and curiosity.
- 2) Designed as Collaborative tool, to integrate data coming from web on the basis of people real traditions.
- 3) Innovative navigation, retrieval, and visualization tools; the portal has been designed to easily guide users in the discovery of the recipes of their interest through the use of a hierarchical filtering paradigm of the relevant information.
- 4) Testbed for tools and algorithms for content based retrieval (text and images).

A. Types of Information

CookIT portal stores multimedia information describing, in Italian language, the traditional recipes handed down from generation to generation. The portal has been implemented as a collaborative tool, to offer the users the possibility to enter their own variants of the recipes.

Two different types of information, to be used for search and visualization purposes, have been identified:

- 1) *Recipes*: They are the kernel of the portal and are defined by title, description, ingredients, variants, place, date, tags ..., integrated with images and videos.
- 2) *Ingredients/nutrients*: Data are taken from or linked to authoritative ingredients/nutrient databases, including open linked data. They consist of ingredient name (and its variations), nutrients, colors, calories, class, ...

B. Analysis and Design

The portal has been designed to store traditional Italian recipes. In order to identify the recipes and foods to be inserted, and populate the portal, the following steps have been identified and implemented:

- Choice of the recipes: People of different ages, social conditions, and Italian regions were interviewed to identify the recipes they consider as part of their own tradition and culture;
- Identification of recipe schema: In order to organize the data with the search results in mind, an analysis of the markup schema specific for foods and recipes has been performed. The most used standard, supported by Google, Bing, and Yahoo (allowing rich snippets) is schema.org/recipe [28]-[30];
- Identification of Italian websites that implement schema.org/recipe microdata, thus allowing to automate as much as possible the process of extracting data from sites, structured by ingredients (and their quantity), URL, directions, steps, images, etc.;

- Design and implementation of the database structure for recipes and ingredients/nutrients: The step should consider recipe standards, metadata information, and popular cooking and recipes web sites as Wikibooks Cookbook [15];
- Identification of recipes and insertion of data in the recipe database, in a semi-automatic way. According to different websites, besides ingredients and directions, calories, portions and even step-by-step images can be used;
- Identification of food and nutrients databases and/or linked open data: Different nutrients and food databases, including open linked data, are considered, related to the kinds of information treated and the languages in which they are available. Data from, for example, FAO Agrovoc (a Multilingual thesaurus) [31], Food Composition Database for Epidemiological Studies in Italy in Italian [32], or USDA Nutrient database (United States Department of Agriculture) [33], can be used to generate and populate ingredients/nutrients schema.

C. Multimodal Navigation and Retrieval

Multimodal means of navigation and retrieval have been designed and implemented. CookIT offers a standard retrieval interface that allows differentiate searches, and a second step to reduce the outcomes to more focused results.

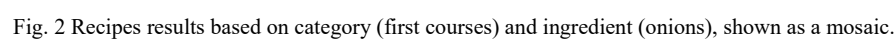
The user can query the database by means of different concepts:

- Food name: "Pasta alla Amatriciana"
- Category: "first courses", "bread and pizza", "quiche" ...
- Ingredients: "artichokes", "peeled almonds or pistachios"
- Place/around me: "near Rome", "Alps mountain"
- Date: "Christmas", "Easter" etc.
- Color: "red", "violet", "white"
- Nutrients: "vitamin A", "proteins" ...
- Special diets/allergens: "vegan", "allergy to crustaceans"
- Calories: "low-cal", "under 100 cal"
- TAG: "easy recipe", "kids"
- ...

The system retrieves the recipes integrating a combination of methods: tags, co-occurrence of recipe-related words, localization etc. and offers the results shown as a list, as a mosaic, as points in a map, as points in a calendar, as a graph in which users choose the axes, e.g. category vs calories.

Fig. 1 shows the search by Region: a detail of Genoa area. The graphic signs in red indicate the category of the recipe (e.g. first courses, unique dishes, street food, appetizers, side dishes, bread/pizza ...), while the round icons in blue indicate to zoom, because more items are in the same position. The retrieval results can be narrowed by applying a further filter chosen among Category, Tag or Ingredients.

Fig. 2 shows a two-step retrieval result: first courses (category search) containing onions (ingredient search), displayed as a mosaic.



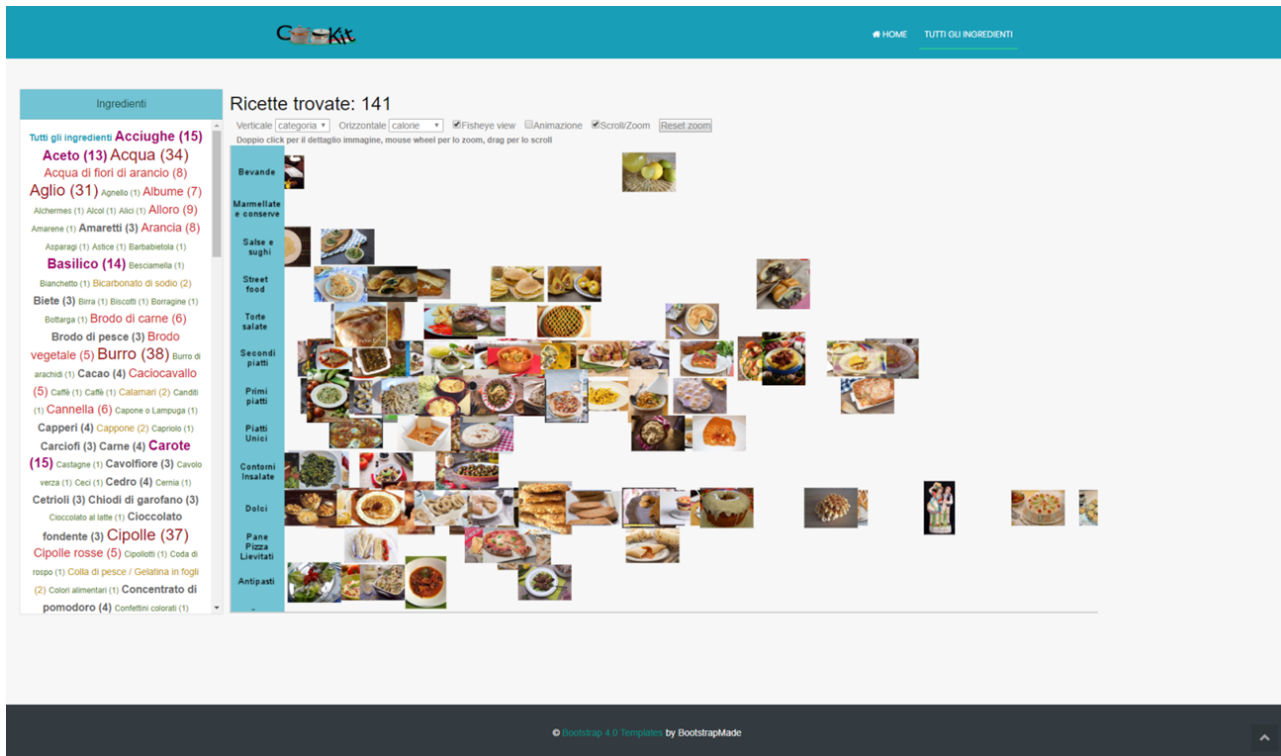


Fig. 3 Recipes shown as a graph, Calories vs Category

Fig. 3 shows the recipes on a graph, in which the horizontal axis represents the Calories, while the vertical axis is the Category. Other possible choices for the axis are regions or colors. On the left the list of ingredients can be used to narrow the search. The recipes in the same intersection (Calories/Category) can be browsed by moving the mouse over, and a double click opens the detail view.

Recipe detail view contains:

- Some basic information as ingredients, directions, steps;
- One or more images and videos;
- A brief description of the recipe and, if available, the history of the dish;
- URLs of the original recipe and the image;
- Tags;
- Suggestions of similar recipes;
- Suggestions of similar images.

Fig. 4 shows the long page of "Agnolotti del Plin", a type of ravioli, with a filling made with two types of meat and some vegetables. It is a typical Piedmontese recipe.

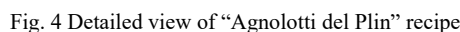
The portal has been implemented using Bootstrap Framework 4.0 and PHP language, storing data into a MySQL database. Both the interface language and the language of the recipes are currently limited to Italian, but the structure allows a possible double language, adding English. At the moment, 140 recipes have been added to the portal, and about 450 traditional Italian recipes have been identified and are being edited.

IV. FUTURE WORKS AND CONCLUSIONS

In the paper, CookIT has been presented, a portal for the preservation and dissemination of Italian traditional recipes. CookIT portal has been designed as a dissemination website to promote the culture of the Italian traditional cuisine, new styles of data visualization are under construction, such as the use of color of ingredients and a MyPlate like visualization mode [34].

CookIT portal has been designed also (and perhaps mainly) as a way to collect knowledge in forms of recipes and images about the many differences in the Italian regional cuisine. It has been projected to be a research tool on which to experiment with new visualization and search algorithms for content, images and texts.

Clustering of recipes on the basis of ingredients, courses, colors, as well as suggestions of similar food for ingredients, occasion, has been planned. The possibility of building healthy and balanced menus is being studied with the help of expert nutritionists.



175

- [25] Le terre dei Trabocchi – Abruzzo (the lands of the trabocchi - abruzzo) <http://www.visitterredeitrabocchi.it/> (last accessed on 10/10/2018)
- [26] Krug, Steve. Don't make me think!: a common sense approach to Web usability. Pearson Education, 2nd Edition, 2005.
- [27] Rosenfeld, Louis, and Peter Morville. Information architecture for the world wide web. " O'Reilly Media, Inc.", 2004.
- [28] Goel, K., Guha, R. V., & Hansson, O. (2009). Introducing rich snippets. Google Webmaster Central Blog.
- [29] Ronallo, Jason. "HTML5 Microdata and Schema. org." Code4Lib Journal 16 (2012).
- [30] Recipe schema on schema.org <https://schema.org/recipe> (last accessed on 10/10/2018)
- [31] FAO Agrovoc <http://aims.fao.org/vest-registry/vocabularies/agrovoc> (last accessed on 10/10/2018)
- [32] Food Composition Database for Epidemiological Studies in Italy <http://www.bda-ico.it/> (last accessed on 10/10/2018)
- [33] USDA Nutrient database <https://ndb.nal.usda.gov/ndb> (last accessed on 10/10/2018)
- [34] MyPlate <https://www.choosemyplate.gov/> (last accessed on 10/10/2018)
- [35] Mori, S., Sasada, T., Yamakata, Y., & Yoshino, K. A machine learning approach to recipe text processing. In: Proceedings of the 1st Cooking with Computer Workshop. 2012. p. 29-34.
- [36] Ciocca, G., Napoletano P., and Schettini R., "Food Recognition: A New Dataset, Experiments, and Results," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 3, pp. 588-598, 2017.
- [37] Ciocca G., Napoletano P., Schettini R., CNN-based features for retrieval and classification of food images, Computer Vision and Image Understanding, in press. <https://doi.org/10.1016/j.cviu.2018.09.001>.

Maria Teresa Artese took her degree in Computer Science at the University of Studies of Milan in 1990. She has been working at the CNR since 2000. Now she works at IMATI – CNR Unit of Milan. Dr. Artese major areas of work are functional analysis and software development, technical support, database structuring and development, dynamic web database sites, design and implementation. Recently she has focused his research on multimedia information systems development and data online management. On these topics she has several national and international publications, and she has been working in national and international research projects.

Gianluigi Ciocca took his degree (Laurea) in Computer Science at the University of Milan in 1998, and since then he has been a fellow at the Institute of Multimedia Information Technologies of the Italian National Research Council, where his research has focused on the development of systems for the management of image and video databases and the development of new methodologies and algorithms for automatic indexing. He is currently assistant professor in computer science at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano-Bicocca, working on video analysis and abstraction.

Isabella Gagliardi took her degree in Physics at the University of Studies of Milan in 1985. She has been working at the CNR since 1986. Now she works at IMATI – CNR Unit of Milan. Dr. Gagliardi major areas of research include Hypermedia Information Retrieval models and methodologies, automatic generation of hypertextual links between text-text, text-image, and audio-audio, dynamic web based database design and implementation, and clustering algorithms. Recently he has focused his research on the development of multimedia information systems available on the web and development of participative online platform.