Estimating Nutrition Values for Internet Recipes

Manuel Müller[1] Stefanie Mika[1] Morgan Harvey[1] David Elsweiler[2]

[1]Chair of Artificial Intelligence

University of Erlangen-Nuremberg, Erlangen-Germany

[2] Institute for Information and Media, Language and Culture

University of Regensburg, Germany

e-mail: ichbin@dermanuel.com, stefanie.mika@cs.fau.de, morgan@derharvey.de, david@elsweiler.co.uk

***Abstract*—**
To utilise the vast recipe databases on the Internet in intelligent nutritional assistance or recommender systems, accurate nutritional data for recipes is needed. Unfortunately, most recipes have no such data or have data of suspect quality. In this demo we present a system that automatically calculates the nutritional value of recipes sourced from the Internet. This is a challenging problem for several reasons, including lack of formulaic structure in ingredient descriptions, ingredient synonymy, brand names, and unspecific quantities being assigned. Our results show that our system can generate nutritional values within a 10% error bound of human assessors for calorie, protein and carbohydrate values. Based on our findings this is smaller than the bound between multiple human assessors.

***Index Terms*—Lifestyle, Health, Prevention, Recommender Systems, Demo**

I. Introduction and Motivation

Poor dietary habits are a major cause of global health problems in the modern world. The World Health Organisation (WHO) predicts that the number of obese adults worldwide will reach 2.3 billion by 2015 [1]. A large body of evidence suggests that health problems like diabetes or heart problems can be prevented and sometimes even reversed through good nutrition [2]. But the problem remains that people are very poor at judging the healthiness of their diet [3] and, even recognising a problem, lack the knowledge of nutritional principles needed to change their diet to a healthier one [4]. Although much of the information needed is available in books, magazines, television programmes and the Internet, people lack the knowledge, time and motivation required to exploit these resources.

Different technological solutions to help asses and improve diets have been proposed as a solution to this problem [5]. One is to design automated systems to plan or provide meal recommendations for individuals based on personal nutritional needs, tastes, cooking skills and lifestyle e.g. [6]. Other efforts have tried to better understand the user's tastes to improve recommendations [7]. Our work aims to build on these initial projects by developing systems that not only recommend recipes according to personal preferences, but combine recipes into dietary plans conforming to WHO nutritional guidelines [8] and user activity profiles derived from sensor technology.

A necessary pre-requisite to building any of the systems described above and implementing them in practical situations is to have appropriate nutritional information for recipes in the database.

II. The System

The system we present automatically calculates the nutritional content of recipes sourced from the Internet using the official nutritional table of the German ministry for nutrition, agriculture and consumer protection which consists of the nutritional values (energy, fat, protein, etc) for over 15,000 items [1]. This is the largest available German database, is reliably sourced and covers a very broad range of ingredients and is appropriate for our intended users as well as our recipe collection. Our collection consists of 23,500 recipes containing a total of 39,500 different listed ingredients obtained from chefkoch.de, a popular German website with a very large and varied collection of recipes.

A. System Architecture

An overview of the main components of the system and how they work together is shown in Fig 1. For each ingredient the raw description is taken from the source text and separated into the *amount* and *ingredient description* (1). Both parts are processed separately in Fig 1:right and Fig 1:left. The output from these components is combined to calculate the nutritional property for the ingredient(10). The values for all ingredients are summed to calculate the nutritional properties for the complete recipe.

To match the ingredient to an appropriate BLS entry, its description is preprocessed (2) by removing punctuation and conversion to lower case. At this stage any appropriate description conversion rules, taking the form of "white fish → haddock", are applied to counter the problem of synonymity and lack of specificity. The first word in the description or rule output is isolated and stemmed and the database is queried using both the original and stemmed versions (3). This combined list of matches is then ranked by a weighted ranking model (4), which we trained from a collection of manually provided assignments from human assessors. The ranking function deals with specificity problems reducing several potential matches to the top-ranked ingredient to be used.

To determine an appropriate weight in grams for the ingredient (Fig 1:left), the raw description of the amount is first split into quantity and unit (5). The unit is, if necessary, converted according to rules like "1 potato → 60g". This rule list was generated by choosing the most frequent unknown units and obtaining the correct conversion ratio

[1] Bundeslebensmittelschlüssel (BLS) http://www.bls.nvs2.de/

PervasiveHealth 2012, May 21-24, San Diego, United States

DOI 10.4108/icst.pervasivehealth.2012.248682

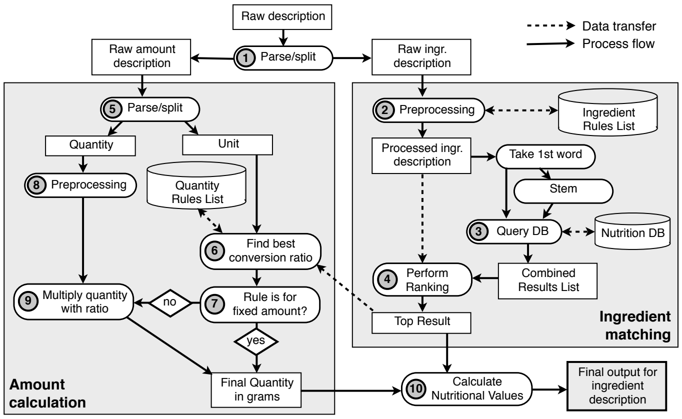

Fig. 1
Diagram of system

Recipe Description		Best Database Match	
200g	pasta (tagliatelle), dried	*200g*	raw egg pasta
100g	ham, smoked bacon	*100g*	pork belly bacon, raw
200g	mushroom	*200g*	raw mushrooms
200ml	cream	*200g*	cream
2 tbsp	squeezed juice from lemon	*30g*	lemon juice
pinch	parsley, dried	*5g*	dried parsley
2	garlic cloves	*10g*	garlic
2 tbsp	butter	*20g*	butter
250ml	bouillon	*2g*	granulated meat stock

	Total	Per portion	Per 100g	% GDA
Energy (kcal)	2551	850.3	332.6	34
Protein (g)	59.6	19.9	7.8	36.2
Fat (g)	190.4	63.5	25.0	66.8
Carbohyrates (g)	153.3	51.1	20.1	17
Fibres (g)	12.1	4.0	1.6	16.7

Fig. 2
Screenshot of demo system

based on the USDA (United States Department of Agriculture) food database[2]. To determine which rule should be applied, the matched ingredient description might be needed (dashed arrow linking the two main components in Figure 1). If the chosen rule is for a fixed amount without any specific quantity (7), e.g. a "dash of cream", then the final quantity in grams is returned. Otherwise the conversion ratio is multiplied by the specified quantity (9). Once the system has selected a single ingredient and a final amount in grams, the complete nutritional properties of the item can be calculated.

B. Demo System

Figure 2 shows an example screenshot of the output from the system which can be rapidly generated on-the-fly from any chefkoch source recipe. In the top table the original recipe data from chefkoch is shown on the left and the system's choices on the right. Notice that the system is able to recognise and convert completely non-standard measurements into grams and is able to find a good match for each ingredient in the recipe in the BLS database. The bottom table displays the estimated nutritional values for this recipe in difference ways: the total values, the values per portion, per 100 grams and finally as a percentage of the daily recommended intake for a man. The last column also gives users a useful overview at a glance of how healthy this particular dish is by using the EU-standard traffic light system. In this case indicating that this recipe contains a large percentage of fat, is a good source of protein, however not such a particularly good source of fibres. The system is able to find a match for all ingredients in 91.1% of recipes (<1% of the recipes have more than 1 ingredient missing). This justifies exploiting the long-tailed distribution in our rule creation process. In comparison to 6 human researchers led by a nutritionist our system came within a 10% error bound for calorie, protein and carbohydrate values, with this bound being smaller than the bound between multiple human assessors.

[2] http://ndb.nal.usda.gov/ndb/foods/list

III. Conclusions and Future Work

We intend to use the presented system as a foundation for a larger recommender system that can automatically generate healthy menu plans for one or several weeks in advance based on the user's tastes and profile, accounting for features such as novelty and diversity. For this purpose we are working with a nutritionist and have collected user ratings from 160 users for chefkoch recipes together with reasons for the ratings. We plan to use this system to understand if the nutritional content of recipes (among other influences, e.g. preparation time, preparation ease) affects how appealing it is to users.

References

[1] "World health organization: Chronic disease information sheet http://www.who.int/mediacentre/factsheets/fs311/en/index.html (accessed feb 14th 2012)," .

[2] Diabetes Prevention Program Research Group, "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin," *New England Journal of Medicine*, vol. 346, no. 6, pp. 393–403, 2002.

[3] G. Johansson, A. Wikman, A. M. Ahrn, G. Hallmans, and I. Johansson, "Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living," *Public Health Nutrition*, vol. 4, no. 4, pp. 919–27, 2001.

[4] Joanne F. Guthrie, Brenda M. Derby, and Alan S. Levy, *America's Eating Habits: Changes and ConsequencesAgriculture Information Bulletin No. (AIB750)*, chapter What People Know and Do Not Know About Nutrition, pp. 243–280, USDA, 1999.

[5] F.E. Thompson, A.F. Subar, C.M. Loria, J.L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *Journal of the American Die- tetic Association*, vol. 110, pp. 48–51, 2010.

[6] K. Hammond, "Chef: A model of case-based planning," in *Proceedings of the National Conference on AI*, 1986.

[7] J. Freyne and S. Berkovsky, "Intelligent food planning: personalized recipe recommendation," in *Proceedings of the 15th international conference on Intelligent user interfaces*, New York, NY, USA, 2010, IUI '10, pp. 321–324, ACM.

[8] Nutrition Joint WHO/FAO Expert Consultation on Diet, the Prevention of Chronic Diseases, and World Health Organization., *Diet, nutrition and the prevention of chronic diseases: report of a Joint WHO/FAO Expert Consultation*, World Health Organization, 2003.