

NUTRITION SYSTEM DEMONSTRATION

Mandy Korpusik, Rachael Naphtal, Nicole Schmidt, Scott Cyphers, and James Glass

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA
{korpusik, rnaphtal, nicole16, cyphers, glass}@csail.mit.edu

ABSTRACT

This demonstration is an initial prototype of a larger effort to create a nutrition dialogue system that automatically extracts food concepts from a user’s spoken meal description. First, the user’s spoken input is recognized by a speech recognizer. Then, the language understanding component uses a semi-Markov conditional random field (CRF) to segment and assign semantic tags to each token in the recognized food diary, and a second CRF associates foods with their corresponding attributes. Finally, the segmented food concepts are shown in matrix form in a table along with potential matches to the USDA and other on-line nutritional databases.

1. INTRODUCTION

Existing approaches for the prevention and treatment of obesity are hampered by the lack of accurate, low-burden methods for self-assessment of food intake, especially for hard-to-reach, low-literate populations [1, 2]. For this reason, we are exploring whether speech understanding and dialogue technology can be used for efficient self-assessment of energy and nutrient intake. We are interested in studying whether speech can lower user burden compared to existing self-assessment methods, whether spoken descriptions of food intake can accurately quantify caloric and nutrient intake, and whether dialogue, perhaps in conjunction with other modalities, can efficiently and effectively ascertain and clarify food properties.

This prototype demonstrates our current progress in the extraction of food concepts from a user’s spoken meal description (e.g., extracting “a bowl of Quaker Oats” from the food log “for breakfast today I had a bowl of Quaker Oats and a chocolate chip muffin”), as well as retrieving relevant matches for food items from a nutritional database.

2. APPLICATION DESIGN

2.1. User Interface

The current interface (see Fig. 1) is configured with a single button to initiate speech recording. The recognized words are displayed on the web-page, along with any associated semantic labels determined by the CRF. A table matrix is shown, with rows corresponding to individual food items, and

columns corresponding to the Brand, Quantity, and Description properties. An image associated with the top result from the nutritional database is displayed for each food item. Finally, for each food, we display the top ten matching items from the USDA and other on-line databases.

The nutrition system uses the SpeechRecognition API in the W3C Web Speech API Specification for audio capture and recognition in the web browser. This API supports continuous listening as well as user-specified start and stop capture, specification of language, whether interim results are sent, and the maximum N for N-best results. For user security reasons, the API requires permission from the user before initiating capture, but will remember the permission if the web application uses https. To use the interface, the developer need only create a SpeechRecognition object in JavaScript, configure it, and register an event handler for the recognition results. At this time, the API is only supported on Google Chrome.

2.2. Database

We examined potential sources of nutrition information, including the USDA and other on-line databases. The USDA maintains a free nutritional database of over 8,000 foods, including basic food items such as fruits, vegetables, eggs, milk, and bread. However, it is missing many common products that consumers would purchase, so we supplemented it using food products from on-line databases. We used the search APIs to query the databases for particular food items, and have been investigating different methods to search for particular foods. Ongoing work involves narrowing down the database hits to a few relevant choices and translating the user-specified quantities into database quantities.

2.3. Data Collection

We deployed a three phase task on AMT to crowdsource our data collection and annotation. The first phase involved the collection of food diaries, where we prompted Turkers to write meal descriptions as if describing them orally. The diaries were tokenized and used as input for the second phase, where we asked users to label food items within the diaries. The third phase combined the food diaries with their food labels and prompted Turkers to label the concepts associated

Food Log

Start Listening

for breakfast today I had **a bowl** of **Quaker Oats** and **a chocolate chip muffin**

Image	Food	Quantity	Brand	Description	USDA Hits	Semantic3 Hits
	muffin	a		chocolate chip	<ul style="list-style-type: none"> Muffins, plain, prepared from recipe, made with low fat (2%) milk Muffins, blueberry, commercially prepared (Includes mini-muffins) Muffins, blueberry, dry mix Muffins, blueberry, toaster-type Muffins, blueberry, prepared from recipe, made with low fat (2%) milk Muffins, corn, commercially prepared Muffins, corn, dry mix, prepared Muffins, corn, toaster-type Muffins, corn, prepared from recipe, made with low fat (2%) milk Muffins, oat bran 	<ul style="list-style-type: none"> Entenmann's Little Bites 5 ct Chocolate Chip Muffins 8.25 oz, Brand: Entenmann's Otis Spunkmeyer® Muffin Variety - 15/4 oz., Brand: Otis Spunkmeyer Betty Crocker Chocolate Chip Muffin & Quick Bread Mix, 14.75 oz, Brand: Betty Crocker Martha White Muffin Mix, Chocolate Chocolate Chip, 7.4-Ounce Packages (Pack of 12), Brand: Martha White Cherrybrook Kitchen Chocolate Chip Muffin Mix, Brand: Cherry Brook Kitchens Krusteaz Muffin Mix Chocolate Chip, 18.25 Oz, Brand: Krusteaz Hostess Mini Muffins - Chocolate Chip, Brand: Hostess Martha White Chocolate Chip Muffin Mix, 7-Ounce (Pack of 12), Brand: Martha White Nuffins Gluten Free Muffin Mix - Case of (4) 20.6-oz Boxes, Brand: Nuffins Krusteaz Chocolate Muffin Mix, Double Chocolate, 20-Ounce Boxes (Pack of 12), Brand: Krusteaz
	Oats	a bowl	Quaker		<ul style="list-style-type: none"> Cereals ready-to-eat, QUAKER, 100% Natural Granola, Oats, Wheat and Honey Cereals, oats, regular and quick, not fortified, dry Cereals, oats, regular and quick, unenriched, cooked with water (includes boiling and microwaving), without salt Cereals, oats, instant, fortified, plain, dry Cereals, oats, instant, fortified, plain, prepared with water (boiling water added or microwaved) Cereals, oats, instant, fortified, with cinnamon and spice, dry Cereals, oats, instant, fortified, with cinnamon and spice, prepared with water Cereals, oats, instant, fortified, with raisins and spice, dry Cereals, oats, instant, fortified, with raisins and spice, prepared with water Cereals, oats, regular and quick and instant, unenriched, cooked with water (includes boiling and microwaving), with salt 	<ul style="list-style-type: none"> Quaker Oats Old Fashioned Oatmeal, 18-Ounce Canisters (Pack of 6), Brand: Quaker Quaker Old Fashioned Oats 42 oz, Brand: Quaker Quaker Natural Granola Oats, Honey, Raisins and Almonds - Two 34.5oz Bags, Brand: Quaker Quaker oats, old fashioned, 2 5 lb. bags, 100+ servings 10-lb, Brand: Quaker Quaker Hot Oat Bran Hot Cereal, 16-Ounce Boxes (Pack of 6), Brand: Quaker Quaker Oat & Yogurt Strawberry Sandwich Biscuits, 1.34 oz, 6 count, Brand: Quaker Quaker Oats - 2/4.5 lb. bags, Brand: Quaker Quaker Oats Oats Quick, 42 oz, Brand: Quaker Quaker Oats Quick Oatmeal, 18-Ounce Packages (Pack of 6), Brand: Quaker Quaker Oats Quick 1 Minute Oatmeal 42 oz, Brand: Quaker

Fig. 1. A prototype of the nutrition system with an example food diary.

with a particular food item.

We collected and labeled approximately 6,000 food diaries for breakfast, lunch, dinner, and snack, which we used to train our language understanding models. We used the AMT label for a token if any one out of five Turkers labeled the token as a food item or property. Every tenth query was added to the test set, while all other queries are part of the training data. Turkers have high agreement when labeling foods and quantities, but more conflicts among brands and descriptions.

2.4. Language Understanding

The language understanding component of the nutrition system has two phases: semantically tagging food concepts and properties in a meal description, and assigning attributes to the correct food items. To accomplish the semantic tagging task, we utilized a variation of the standard CRF model, a semi-Markov conditional random field (semi-CRF). Rather than assigning an output label to each token, a semi-CRF assigns output labels to token *segments* [3]. In the second phase, we investigated three approaches for determining which attributes were associated with which foods: a Markov model, transformation-based learning, and a CRF. The CRF model performed best, achieving a token-level accuracy of 97.22%

and a phrase-level F1 score of 87.13 [4].

3. ACKNOWLEDGEMENTS

Jennifer Drexler, Jeffrey Sun, Daniel Li, and Panupong Pasupat helped with the development of early versions of the data collection tasks for AMT. This research was sponsored in part by a grant from Quanta Computing, Inc., and by the NIH.

4. REFERENCES

- [1] Y. Wang and M. Beydoun, "The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: A systematic review and meta-regression analysis," *Epidemiologic reviews*, vol. 29, no. 1, pp. 6–28, 2007.
- [2] World Health Organization, *Obesity: Preventing and Managing the Global Epidemic*, Number 894. World Health Organization, 2000.
- [3] S. Sarawagi and W. Cohen, "Semi-markov conditional random fields for information extraction," in *Proc. NIPS*, 2004, pp. 1185–1192.
- [4] M. Korpusik, N. Schmidt, J. Drexler, S. Cyphers, and J. Glass, "Data collection and language understanding of food descriptions," in *Proc. SLT*, 2014.