

# Jamie LOD-iver: Enriching Historical Recipes with Linked Open Data

Melvin Wevers, *KNAW Humanities Cluster, DHLab*

Marieke van Erp, *KNAW Humanities Cluster, DHLab*

Hugo Huurdeman, *University of Amsterdam, University of Oslo*

Richard Zijdemans, *International Institute for Social History, University of Stirling*

## Introduction

What we eat has been an important part of our cultural identity.<sup>1</sup> There are monthly recipe magazines,<sup>2</sup> cooking shows,<sup>3</sup> and in Amsterdam alone, twenty-one culinary festivals are scheduled for 2018.<sup>4</sup> What we eat can be distilled from recipes, which are easily accessible for contemporary research, as recipes on food websites are marked up with hRecipe or schema.org.<sup>5</sup> Unfortunately, this is not the case for historical recipes, often available as part of digitized archives. This means that currently we are unable to answer important questions on how food culture has changed over time. Are we using more sugar in home cooking today? To what extent has migration affected eating habits?

In this poster proposal, we present a method for extracting and enriching over 4,000 recipes from digitized historical newspapers (*Het Parool, Trouw, De Volkskrant, and NRC Handelsblad*) published between 1950 and 1995 via Delpher.<sup>6</sup>

## Method: Extracting and Enriching Newspaper Recipes

To extract and enrich recipes from newspapers, we devised the workflow depicted in Figure 1.

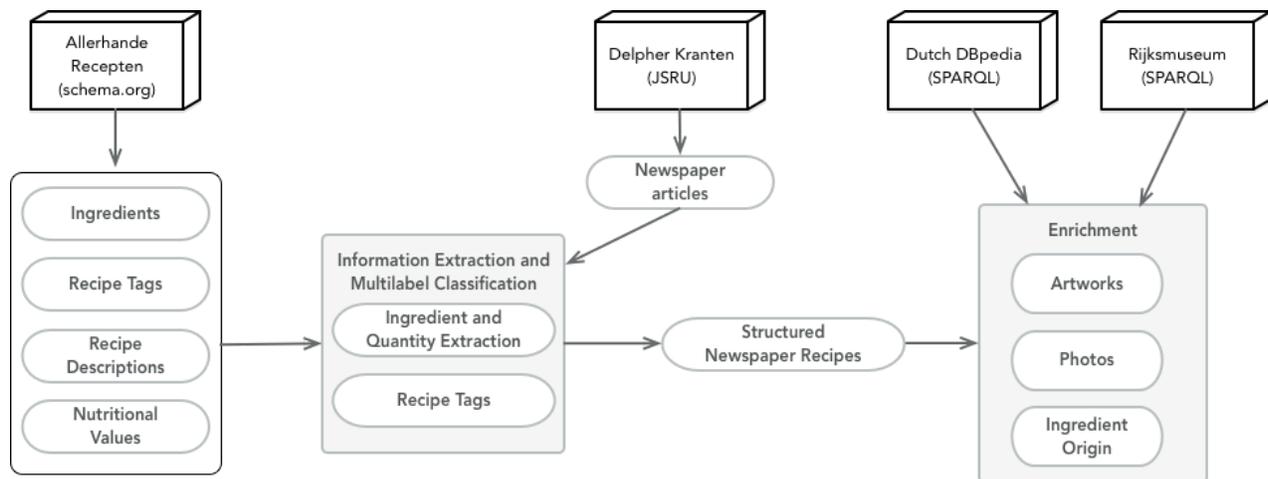


Figure 1: Newspaper recipe extraction & enrichment workflow

<sup>1</sup> Whatmore, Sarah, and Lorraine Thorne. "Nourishing Networks: Alternative Geographies of Food." In *Globalising Food: Agrarian Questions and Global Restructuring*, edited by David Goodman and Michael Watts, 211–24. London: Routledge, 1997; Wilson, Thomas M., ed. *Food, Drink and Identity in Europe*. European Studies 22. Amsterdam: Rodopi, 2006; Otterloo, Anneke H. van. *Eten en eetlust in Nederland, 1840-1990: een historisch-sociologische studie*. Amsterdam: Bert Bakker, 1990.

<sup>2</sup> <https://deliciousmagazine.nl/>; <https://www.foodiesmagazine.nl/>

<sup>3</sup> <https://heelhollandbakt.omroepmax.nl/>

<sup>4</sup> <https://www.iamsterdam.com/en/see-and-do/whats-on/festivals/overview-culinary-festivals-and-events>

<sup>5</sup> <http://microformats.org/wiki/hrecipe>; <http://schema.org/Recipe>

<sup>6</sup> Because of higher OCR quality for more recent newspapers, we selected these newspapers. For more on these papers, see: <https://www.kb.nl/nieuws/2017/belangrijke-naoorlogse-kranten-digitaal-beschikbaar>

We first bootstrapped 16K recipes from the *Allerhande* recipe website to obtain lists of ingredients and recipes belonging to specific to specific tags such as 'vegetarian', 'Christmas', and 'Italian'. The recipe descriptions and their associated tags were used to train a multilabel classifier, needed to assign such tags to the recipes from the newspapers (that were selected via seed words and metadata). Next, we applied a rule-based tagger based on *Allerhande's* ingredients list to extract ingredients and their quantities from the recipes. The recipes were further enriched with photographs from Rijks Museum and Dutch DBpedia using SPARQL endpoints. From the latter source, we also extracted location information associated with the ingredients to enable a geographical mapping of recipes.

### Analysing Recipes

By enriching the recipe information, we open up the possibility for easy exploration, comparison and quantitative analyses. Additional text analysis such as Topic Modeling and Word Embedding can guide qualitative studies of changing food culture. Figure 2 visualizes a selection of the newly enriched recipes in a network diagram, organised by their ingredients.



Figure 2: Recipes organised by their ingredients in a network diagram, using images from DBpedia

We are in the process of extracting and classifying the nutritional information of the extracted recipes. This presents several hurdles as not all quantities are stated very precisely or have come through correctly after OCR.<sup>7</sup> These enrichments will enable researchers to compare, for example, whether caloric values are related to the season in which the recipes were published (Figure 3).

<sup>7</sup> For example, the page fold in the scan for an India Relish recipe from 1991 made exactly the quantities unreadable: <https://resolver.kb.nl/resolve?urn=KBNRC01:000030263:mpeg21:a0129>

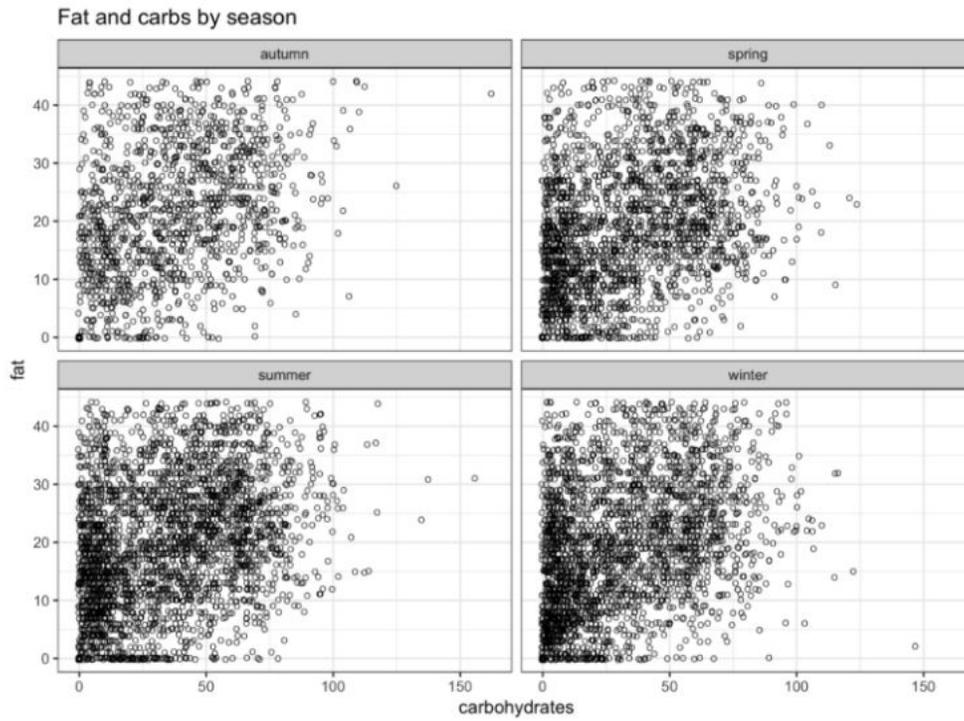


Figure 3: Analysis of carbohydrates and fat content of Allerhande recipes 2000 - 2017

Future work could focus on extracting and enriching recipes from earlier periods and other sources than newspapers, unveiling the evolution of cuisines and the representation of cultural identities in discourse on food.