

Identifying online display of Food Hygiene Rating Scheme ratings: Finding online presence of food businesses

Methods and challenges

The FHRs data does not include an establishment's web presence, and we wanted to avoid having to search manually for this information, to speed up the process both in this project and in any future iterations or re-runs. Therefore this information needed to be brought in from an external source that holds information about businesses, and matched against our data.

There were two main services that could potentially have provided this information: Google Places, and TomTom. Google Places provides data on 'points of interest', including businesses; this is the information about a place that you would see if you searched for it on Google Maps, although the data is used in a variety of ways across different Google and third party applications. TomTom also collects this type of data on places for use in its satellite navigation systems, and also makes this data available to third party developers.

Of the two sources, Google provided by far the better coverage. In a random sample of 150 businesses, Google was able to retrieve details about 112, compared with just 19 for TomTom. Therefore Google Places was used as the method for finding business websites in this project. The cost of retrieving the information from Google Places for a sample of 1500, the size used in this project, is around £60.

Google provides an interface to its data that allows the user to submit a business name and a location (in the form of latitude/longitude co-ordinates), and it will search on that name within a given radius of that location. It then returns data for the corresponding business (or what it considers the best match, if there is more than one). Although these requests need to be submitted individually, a simple program can be written to do this across a sample of businesses, and return a dataset of information about those businesses.

Although the vast majority of establishments will have a presence on Google Places, there will be some that do not, and although Google's algorithm for searching for and retrieving data on the business is very good, and will return the correct data in the most cases, it is not perfect. However it is the best available method for retrieving this kind of information, and quality assurance procedures (detailed below) suggested that the results are of high quality.

This process was applied to our sample of 1500 businesses. Once the website addresses had been retrieved, a further step took place to exclude websites that were social media pages (Facebook, Instagram or Twitter), booking websites (booking.com, Trivago or hotels-247), or the main website, as opposed to a branch-specific page, for one of the 75 largest chains in the UK (e.g. Costa, Starbucks, Dominos).

Results

Full sample

This process of retrieving business URLs returned business websites for 803 businesses – 54% of the total sample. Below summarises the number of businesses that were eliminated at each stage.

- Full sample = 1500
- Matched to a place on Google = 1336 (89%)
- Has a business website = 927 (62%)
- Website is not Facebook, chain level or third party = 803 (54%)

Just under 90% of establishments were matched to a place on Google Places. Of these, 927 had a business website, but after the final exclusion step, there were 803, or just over half of the sample.

Quality assurance

It is not possible to go through the whole sample to check whether this process has worked, and would defeat the purpose of an automated solution. However, a quality check was carried out on a sample of 102 businesses to establish whether the websites were missing because they do not exist, or because they could not be found with our methods, and whether the websites that were returned were the correct ones.

Of these 102 businesses, the process had failed to find a Google Place for 7 of them. A manual check suggested that 4 truly did not exist as a Google Place, 1 was not returned because it was marked as closed, 1 was present but the name recorded in the FHRS data was too different to match, and the other non-retrieval was unexplained. Therefore we can say that coverage of the businesses in the Google Places data is 96%, and our process successfully matches establishments to this data in 98% of cases.

In 32 cases, an establishment had been matched to a place on Google, but there was no business website in the data returned. This was found to be correct in every case; none of these businesses had a website that we could find. Of the 54 websites that were found, 49 were correctly identified. Therefore it is highly unlikely that a website exists and we have not found it, but around 9% of the websites found will not be the correct web address.

This suggests that the process, while not failsafe, has returned a set of business websites that we can be broadly confident are the right ones, and that it is not systematically failing to uncover businesses' online presence. The distribution of business websites across business type and nation is shown in Table 3. Despite the total sample size reducing by nearly half, all business types in all nations are still represented in the sample.

The lack of a business website for almost half the sample raises the question of whether many food businesses are relying on social media for their online presence. As we have excluded such websites from this analysis, our process would be failing to capture a significant aspect of online display if this were the case. We examined a sub-sample of 100 businesses in our sample for which there was no business website. Of these, 37 were found to have a Facebook page. Interestingly, 7 of these Facebook pages were displaying a rating, which is a much higher prevalence of online display than was found among business websites. However the wrong rating was displayed in 4 of these cases. Restrictions on the scraping of social media pages mean that it is difficult to incorporate them in the image matching process developed for this project, however the proportion of 'Facebook only' businesses in the sample is relatively small; around 17% of the sample, or around 256 businesses.

Table 3: Composition of final sample of websites

| | England | Northern Ireland | Wales |
|-----------------------------------|---------|------------------|-------|
| Hotel/bed & breakfast/guest house | 22 | 30 | 43 |
| Pub/bar/nightclub? | 59 | 23 | 61 |
| Restaurant/Cafe/Canteen? | 144 | 127 | 124 |
| Takeaway/sandwich shop? | 71 | 49 | 53 |
| Total | 296 | 229 | 281 |

Options for future iterations

At the start of the project it was not known what proportion of the sample would have an online presence. As it transpires that this figure is quite low, a larger sample of businesses would ensure a final website sample size closer to the target. As around 8% of the links in the sample were broken/inaccessible, a larger sample would also compensate for this loss.

As the only restriction was on business type as it is recorded in the FHRS data, this sample includes establishments that may not be truly in scope, for example workplace canteens, and educational establishments (where the FHRS record has been matched to the company or institution rather than its café or catering facilities specifically). If mandatory display is restricted to consumer-facing businesses in future, then this sample should probably be restricted in a similar way, with manual fine-tuning of an initially random sample.

Implications for policy and practice

The difficulties encountered in finding food businesses' online presence would provide some support for requiring businesses to provide this information, at point of registration or inspection, and for local authorities to submit it with their FHRS return.