International Journal of Information Management 32 (2012) 504–512

Contents lists available at SciVerse ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Using Google Analytics to measure visitor statistics: The case of food composition websites

H. Pakkala[a,*], K. Presser[b], T. Christensen[c]

[a] Division of Welfare and Health Promotion, Department of Lifestyle and Participation, National Institute for Health and Welfare (THL), P.O. Box 30, FI-00271 Helsinki, Finland
[b] Institute of Information Systems, Department of Computer Science, ETH Zurich, Universitätstrasse 6, CH-8092 Zürich, Switzerland
[c] National Food Institute, Division of Nutrition, Technical University of Denmark, Mørkhøj Bygade 19, DK-2860 Søborg, Denmark

ARTICLE INFO

Article history:
Available online 10 May 2012

Keywords:
Internet
Food composition
Web analytics
Visitor statistics

ABSTRACT

Measuring visitor statistics is a core activity for any website provider. However, the analytical methods have so far been quite limited, difficult, expensive, or cumbersome. Google Analytics (GA) offers a free tool for measuring and analysing visitor statistics. GA was tested on three food composition websites (Denmark, Finland, and Switzerland). All the websites had a considerable number of visitors, which seemed to increase with the maturity of the website. The results also suggested that there were a considerable number of potential unreached users in Denmark and particularly in Switzerland, thus suggesting that promotion be increased and search engines be taken into account more during website design. About 15–20% of users visited the website more than nine times and about 20% spent there more than 10 min on the site. Following traffic from referring websites showed that most of the visitors could not be categorised as food or nutrition professionals. Our experience showed that GA was quite easy to use and gave useful and versatile information that can be used to compare different websites and improve the website design. Finally, we would like to encourage other food composition website providers to utilise either GA or another of the similar tools available.

1. Introduction

Measuring website traffic and analysing user navigation is a common procedure for any website provider. The monitored items vary from simple statistics ("How many visitors we have per day?") to a complex and comprehensive analysis of the navigation behaviour of website visitors ("Why do some web shop visitors collect many products in their shopping cart and then quit before checkout?"). This information can be used to fine tune the measured website to provide visitors with more content that they are interested in and to improve navigation – often with an ulterior motive of increased advertisement income or selling more products. Whatever the fundamental motive may be, web analytics is a cornerstone in creating customer satisfaction among website visitors (Croll & Power, 2009; Kaushik, 2010).

Web analytics can be defined as the assessment of a variety of data, including web traffic, web-based transactions, web server performance, usability studies, user-submitted information and related sources to help create a generalised understanding of the online visitor experience. There are two major methods for gathering information for this analysis: page tagging and using web server log files (Croll & Power, 2009; Kaushik, 2010; Peterson, 2004, 2006).

Page tagging is done by placing an identification tag in one or more web pages of the website. When this web page is used the page identifying tag and information about the visitor is sent to counter software (usually outside the website) that collects this information for later analysis. The information gathered can vary from simple web metrics like visitor counts, to monitoring the whole website session of a visitor. There are free tagging services available, providing only limited amounts of information, but also commercial tagging services providing full-scale information suites and analytical tools (Croll & Power, 2009; Kaushik, 2010).

Using the log files of your web server enables the collection of large amounts of visitor events without using external services. This information can be tailored and the nature of the collected information depends on the imagination and the programming skills of the website builders. However, in any website with considerable traffic these log files may be huge, making their analysis extremely cumbersome and time consuming (Jansen, 2006; Peterson, 2004). Moreover, as these log files are often tailor-made, it is not easy to find suitable analytical tools and the information needs to be pre-processed before using them. This may mean that you have to build your own software tools before being able to

* Corresponding author. Tel.: +358 20 610 8593.
E-mail addresses: heikki.pakkala@thl.fi (H. Pakkala), karl.presser@inf.ethz.ch (K. Presser), tuchr@food.dtu.dk (T. Christensen).

0268-4012/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijinfomgt.2012.04.008

analyse anything. There is also one major challenge with a log file analysis: different websites are often constructed differently and it is quite difficult to compare different websites with each other.

About three years ago Google Inc. launched a free tool called Google Analytics (GA) (Google Inc., 2009). GA uses page tagging and provides a set of versatile analytical tools. Currently, GA is used by over 80% of the commonly used websites that use traffic analysis tools (W3Techs, 2010). In addition, several scientific articles have analysed GA and evaluated its usefulness as a web analytics tool (Bhatnagar, 2009; Fang, 2007; Hasan, Morris, & Probets, 2009; Plaza, 2009a, 2009b, 2011).

Food composition is information on the concentrations of nutrients and nutritionally important components in foods and is used in many different fields of work, especially in public health and nutrition. It is utilised for example dietary guidelines, food labelling and nutrient claims, food legislation, in epidemiological research on relationships between diet and disease, and devising special diets for healthy people with particular needs (e.g. athletes) (Church, 2009; Williamson, 2005). This information is usually presented in the form of food composition tables or databases. Moreover, food composition websites have now been in existence for more than a decade. However, very little has been published about their visitor statistics or other features of user behaviour, with only a few exceptions (Pakkala, Reinivuo, & Ovaskainen, 2006). It is not known if web analytic studies are just not carried out or if the results are just not published. For those unfamiliar with web analytics, it might be easy to assume that it is technically difficult to implement.

The aim of this paper is to demonstrate that web analytics can be rather easy and simple. We tested the use of GA on three food composition websites – in Denmark, Finland, and Switzerland. Moreover, we set out to answer the following questions:

- How are the websites found by users?
- What is the content used by visitors?
- How often do users come back to the website (and how many new users are there)?
- What do we know about users?
- What devices were used to visit our websites?

2. Materials and methods

2.1. Website characteristics

2.1.1. Danish Food Composition website (Fødevaredatabanken)

The Danish Food Composition website has been present on the Internet since 1992. In its present shape, it carries information on 1049 food items with up to 112 nutrient factors. The information is available in Danish and English (Technical University of Denmark, 2010).

2.1.2. Finnish Food Composition website (Fineli)

The Finnish Food Composition website (Fineli) is a well-established website launched as far back as 1999 (Pakkala et al., 2006). Currently, it contains information about 54 nutrient factors and over 2100 food items. The information is available in Finnish, Swedish and English (National Institute for Health & Welfare. Nutrition Unit, 2010).

2.1.3. Swiss Food Composition website (Schweizer Nährwertdatenbank short CH-NWDB)

The Swiss Food Composition website is part of the CH-NWDB, the Swiss Food Composition database. The website was launched in 2007 and was the start of the implementation of a new management system called FoodCASE that incorporates mainly the standards of the European Food Information Resource network (EuroFIR, 2010) and the state of the art in data quality research. The website has not been actively promoted in Switzerland and contains 935 food items, with about 32 nutrient factors per food item (ETH Zurich, 2010).

2.2. Measurement with Google Analytics

The GA measurement tag was placed on each web page of the websites. As the measurement started at slightly different times, a common period was selected for the analyses and it covered a 5-month period from 1 May to 30 November 2010. The measurement was continuous with the exception of a service failure on the Fineli website from 1 to 4 August 2010 and a maintenance break from 6 to 7 November 2010. Promotion campaigns were not carried out during the measurement period.

2.3. Key performance indicators

The key performance indicators used in this study are described in Table 1.

2.4. Data analysis

The primary analysis was done using GA and the result was downloaded. This result was then combined with the results of the other two websites and (if necessary) further analysis was carried with a spreadsheet or with the SPSS statistical software package (IBM, 2011). Most of the analyses were based on the results provided directly by GA.

In the analysis of the most viewed web pages, the 100 most viewed web pages on each website were listed. Then each of the web pages was classified in groups describing the main content of the web page: “Background information”, “Food item”, “Website home page”, “Navigation, search, a set of search results”, “Nutrient compound”, and “Personal food diary”. Then the number of visits was summed together inside each group and scaled with the total number. This was done separately for each of the study websites. Thus, the result was a distribution of visits by content type for each study website.

The search engine keywords were analysed in a similar manner: the 100 most common keywords used when the website was accessed through a search engine were listed. Then, each of the keywords was classified into groups by their subject “Food name”, “Nutrient compound name”, “Generic food composition search term”, and “Website name”. Sometimes there were several keywords referring to different classes and these were classified as “Combination of several groups of terms”. Consequently, the result was a distribution of visits by keyword subjects for each of the study websites.

The procedure was roughly the same with the analysis of the referring websites. First, the top 100 referring websites were listed with the number of visits and time on site. Then, each of these websites was classified by their main interest group e.g. “Food counselling” or “Health and lifestyle” (the complete classification is presented in Table 6). The group was decided subjectively by visiting each of the referring websites. Then the time on site was weighted by the number of visits. In this case, the result was a weighted average of time on site (of those visitors that accessed from referring websites) grouped by the main interest of the referring website (on each study website). The number of visits was used for counting a pooled visit rank: first the main interest groups were ranked by the number of visits separately for each study website; then the ranks from the study websites were pooled. The same grouping of the main interest group was used also for the analysis of the bounce rates in the Fødevaredatabanken and Fineli websites (the CH-NWDB website did not have enough data for the analysis).

Table 1
Key performance indicators used in this study (according to Croll & Power, 2009; Google Inc., 2009; Kaushik, 2010 except the dissemination rate).

Key performance indicator	Description
Absolute unique visitors	The estimated number of people who visited the website
Average time on site	The average of all users' time on the website
Bounce rate	Bounce rate is the percentage of visits that come to a website and leave it without continuing to other subpages. Bounce rate is a measure of visit quality and a high bounce rate generally indicates that site entrance (landing) pages are not relevant to visitors
Depth of visit	The number of page views of one visitor per visit. The depth of visit is a measure of visit quality. A large number of high page views per visit suggests that visitors interact extensively with the website.
Dissemination rate	Percentage. Absolute unique visitors × home country visit rate/population in the country 2010 (Eurostat, 2011)
Home country visit rate	Percentage. Portion of the visits from the same country as the national website (with e.g. for Fineli, it is the visits from Finland)
Keywords	The keywords that were used when the website was accessed via a search engine result page
Landing page	The web page on a website where the visitor "lands" first, i.e. the web pages which the visitor meets first when entering the website
New visitor	A new visitor is a visit by a visitor who has not been recorded previously. A high number of new visitors indicates strong visitor recruitment
New visit rate	Percentage. The share of the new visitors of total visits
Page views	Page views is the total number of pages viewed on the website and is a general measure of how much the website is used
Returning rate	Percentage. The share of returning visitors from total visits
Returning visitor	A returning visitor is a visit by a visitor who has been recorded previously. A high number of return visitors suggests that the website content is engaging enough for visitors to come back
Time on site	The time a visitor spends on site. One way of measuring the visit quality. If visitors sped a long time visiting the website, they may be interacting extensively with it. However, time on site can be misleading because visitors often leave browser windows open when they are not actually viewing or using the website
Traffic source	
Direct traffic	Visits from people who clicked a bookmark to come to the website or who typed the website's URL directly into their browser
Referring sites	Visits from people who clicked a link to the measured website on another website
Search engines	Visits from people who clicked a link to the measured website on a search engine result page
Visits	The number of visits the measured website receives is the most basic measure of how effectively the website is promoted
Visitor loyalty	The number of the repeated visits by returning visitors. Loyal visitors are usually highly engaged with the brand of the website and a high number of multiple visits indicates a good customer and visitor retention

In the comparison of the devices, the key performance indicators were first calculated for each operating system. Then the operating systems were classified according to their main usage. The 'mobile' device categories were Android, BlackBerry, Danger Hiptop, iPhone, iPod, LG, Nintendo Wii, Nokia, Playstation 3, Playstation Portable, Samsung, Sony, SunOS, SymbianOS; iPad was considered a "Tablet"; and FreeBSD, Linux, Macintosh, NetBSD, OpenBSD, UNIX, Windows were considered "Workstations". These groupings were used in the calculation of weighted averages. The number of visits was used as the weighting factor. The operating system categories used for indicating the device were not fully comprehensive as some of the operating systems could be used in several types of devices and thus the results were simplified and indicative.

3. Results

3.1. General information

We found a very large variation in the numbers of visitors between the websites, from an average of 33 visitors per day to 3524 visitors per day. Otherwise, the main characteristics of the website traffic were quite similar between the websites. The users of the CH-NWDB website visited slightly more pages and spent about 20% more time on the website compared to others. Moreover, the bounce rate of the CH-NWDB website was very much lower and the new visit rate was slightly lower than on the other websites. The visits from the home country were quite dominant for each of the three websites. In addition, Fineli has much dissemination rate compared with the other websites. The key performance indicators are presented in Table 2.

Fig. 1 shows information from the GA dashboard – in this example from Fødevaredatabanken. The visit trend curve (at the top of the dashboard) shows the characteristic weekly fluctuation in the numbers: high numbers during the working week and less users during the weekend. Moreover, the number of the visits diminished during holiday seasons. This phenomenon was observed also on the other websites.

3.2. How are the websites found by users?

Table 3 shows the traffic sources of the websites. Direct traffic was the most common way of entering the website on Fødevaredatabanken and especially for CH-NWDB, where it was almost the sole means of access. For the Fødevaredatabanken and Fineli websites, the traffic sources were distributed more evenly. In addition, the number of referring sites was quite low for CH-NWDB.

Table 2
Key performance indicators of the study websites, measuring period 1 May–30 November 2010. Indicators explained in Table 1.

Key performance indicator (see Table 1)	Fødevaredatabanken	Fineli	CH-NWDB
Visits	117 903	754 573	7131
Visits per day	550	3524	33
Absolutely unique visitors	61 575	414 712	3516
Page views	1 050 395	7 446 917	98 682
Pages/visit	8.9	9.9	13.8
Bounce rate	33.5%	33.4%	9.1%
Average time on site	4 min 56 s	4 min 25 s	6 min 5 s
New visit rate	50.5%	52.9%	48.4%
Visits from home country	80.5%	87.7%	83.9%
Dissemination rate	0.9%	6.8%	0.04%

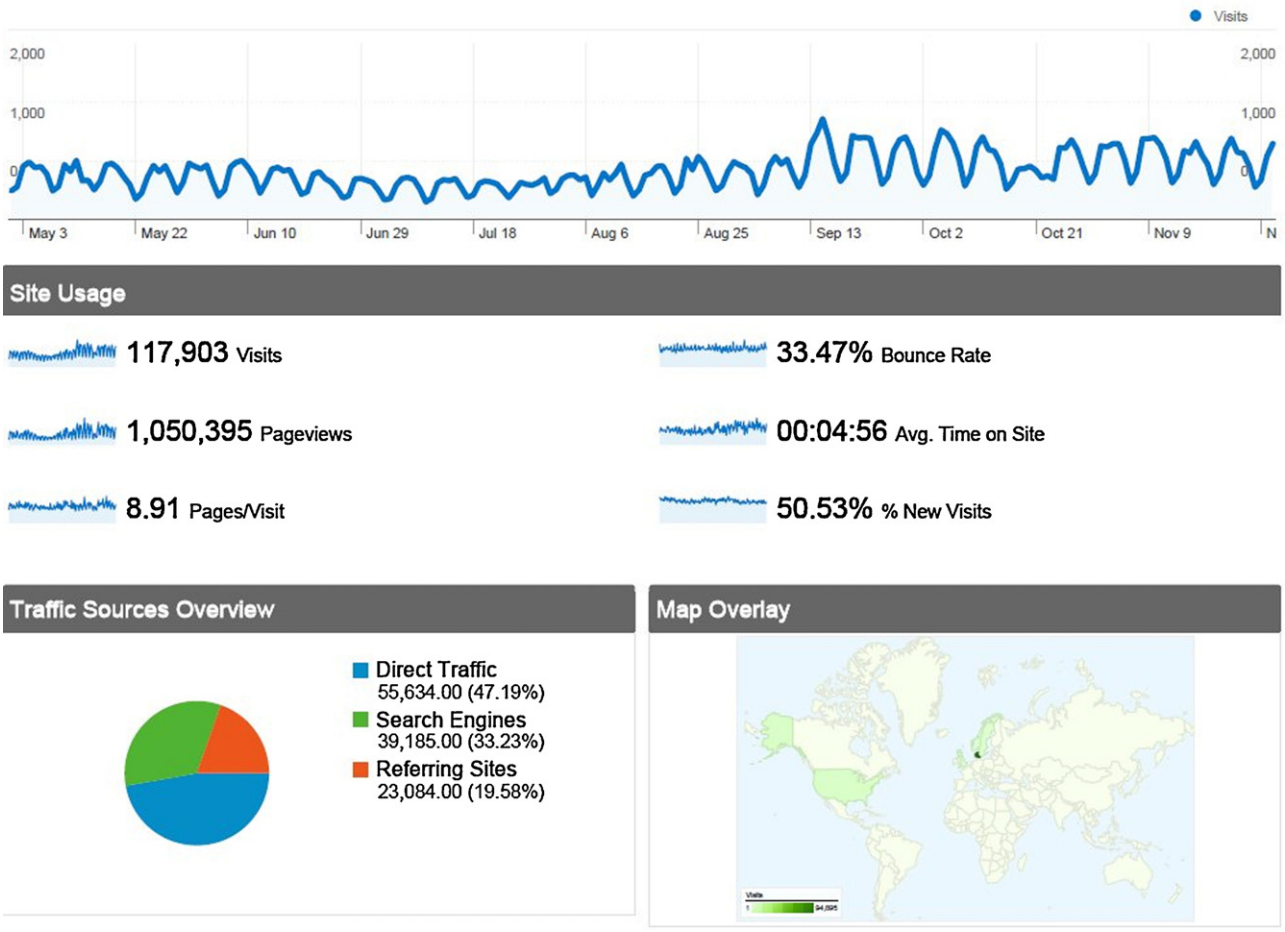

Fig. 1. Example of the information provided by Google Analytics dashboard from the Fødevaredatabanken website (daily data, 1 May–30 November 2010).

Table 3
Traffic sources of the study websites, measuring period 1 May–30 November 2010.

Traffic source (see Table 1)	Fødevaredatabanken	Fineli	CH-NWDB
Search engines	33.2%	49.7%	1.2%
Direct traffic	47.2%	27.7%	93.9%
Referring sites	33.2%	22.5%	4.9%

The keywords used in search engines when the website was accessed via a search results page indicated what kind of information the visitor was looking for. The websites were visited 414 106 times, with the use of 49 514 different keywords in total. The most common 100 keywords were used in total on 55.3% of the visits and Table 4 shows the distribution of the subjects of those top 100 keywords. The name of the website was the most common keyword subject for each of the websites, and on Fødevaredatabanken the generic food composition search terms (such as 'food composition table, 'food table') were used quite often. In general, all keyword subjects were found on the websites except on CH-NWDB, which was accessed via an extremely low amount of keywords (only 85 visits with 15 keywords).

3.3. What content is used by visitors?

The distribution of page views of the 100 most commonly visited web pages was used to indicate the content perused by visitors. Additionally, the average time spent on the web page indicates how interesting the content was to the visitors (Table 5). The results show that web pages relating to the navigation were used more often than other types of web pages. However, when the visitors

Table 4
Distribution of visits by the keyword subjects on the study websites, measuring period 1 May–30 November 2010. Top 100 keywords (CH-NWDB only 15 keywords).

Subject	Fødevaredatabanken $n = 20\,289$ visits	Fineli $n = 208\,686$ visits	CH-NWDB $n = 85$ visits
Name of the website (e.g. 'Fineli')	60.9%	77.2%	68.2%
Food name	7.8%	9.0%	0%
Nutrient compound name	2.5%	6.0%	0%
Generic food composition search term (e.g. 'food composition table')	19.6%	5.8%	31.7%
Combination of several terms (e.g. food name and nutrient compound name)	9.2%	2.1%	0%

Table 5
Distribution of pages viewed and the average time on the page by the content type of the web pages for the study websites, measuring period 1 May–30 November 2010. The 100 most commonly used web pages.

Subject	Fødevaredatabanken $n = 817\,991$ page views		Fineli $n = 5\,709\,700$ page views		CH-NWDB $n = 98\,682$ page views	
	Page views	Average time on page	Page views	Average time on page	Page views	Average time on page
Website homepage	6.9%	24 s	0.4%	30 s	44.4%	21 s
Food	8.2%	1 min 39 s	9.9%	51 s	0	0
Nutrient compound	1.4%	2 min 33 s	0.6%	56 s	0.7%	1 min 49 s
Background information	2.6%	1 min 16 s	0.4%	1 min 2 s	0.3%	1 min 13 s
Navigation, search	80.9%	26 s	56.3%	24 s	54.6%	14 s
Personal food diary	–	–	32.4%	19 s	–	–

found the "real" content (e.g. food or nutrient compound) they were probably looking for, they spent considerably more time on such web pages. A personal food diary was a unique feature on Fineli but there it gained about one third of the page views.

3.4. How often do users come back to our websites?

Visitor loyalty was used to indicate whether the content provided and the layout of the websites were found to be sufficiently adequate and reliable by the visitors that they were willing to return to consult the website again (Fig. 2). About half of the visitors visit the website only once but about 15–20% of the users visited the website more than nine times. On the Fødevaredatabanken and on the Fineli websites, a small portion of visitors visited the website more than 50 times (Fødevaredatabanken 7.1%, Fineli 4.2%) i.e. more than once per week.

3.5. What do we know about the visitors?

The referring sites from which the visitors arrive at the websites give information about the interests of the visitors. We used the average time spent on a page to indicate how important a source our websites were to the different visitor groups. In addition, the number of visits is a rough measure of the popularity of the websites among the different visitors. In total the study websites were referred by quite a lot of external websites (Fødevaredatabanken by 645, Fineli by 1715, CH-NWDB by 14) but visits from the 100 most common referrers constituted about 90% (in CH-NWDB 100%) of all referrers. The visitors who spent most of the time on the study websites seemed to be interested in subjects like food education, home cooking and gardening, and health and lifestyle. Also

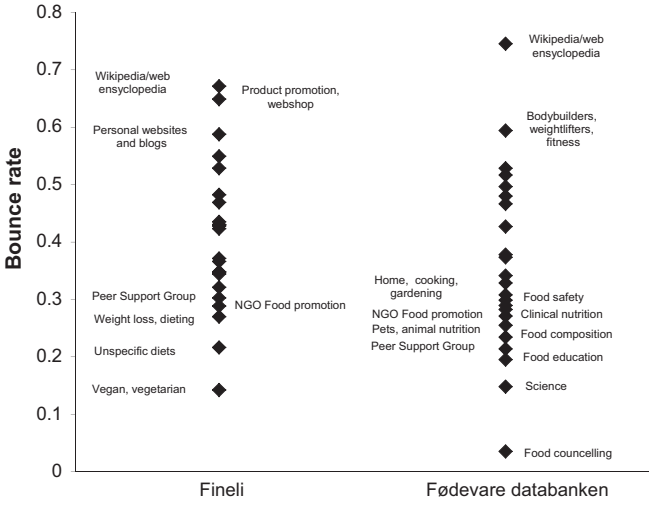

Fig. 3. Bounce rate when arriving from a referring site, grouped by the main interest groups. Fødevaredatabanken and Fineli website, measuring period 1 May–30 November 2010. Top 100 website referrers. The labels are presented with the high and the low bounce rates.

the visitors of product promotion websites and the readers of personal websites or blogs were an important user group. In addition, the study websites gained most traffic from people interested in health and lifestyle or physical exercise, and from the readers of Wikipedia (Wikimedia Foundation Inc., 2011), news and social media (Table 6). Fig. 3 shows the bounce rates among the different visitor groups. The low bounce rates indicate that the website is relevant to the visitor. The lowest bounce rates were found from

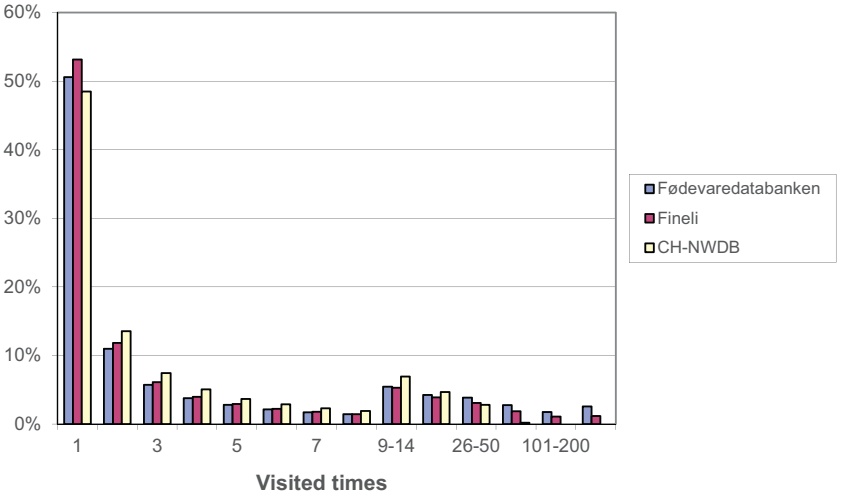

Fig. 2. Distribution of the visitor loyalty for the study websites, measurement period was 1 May–30 November 2010. Top 100 web pages.

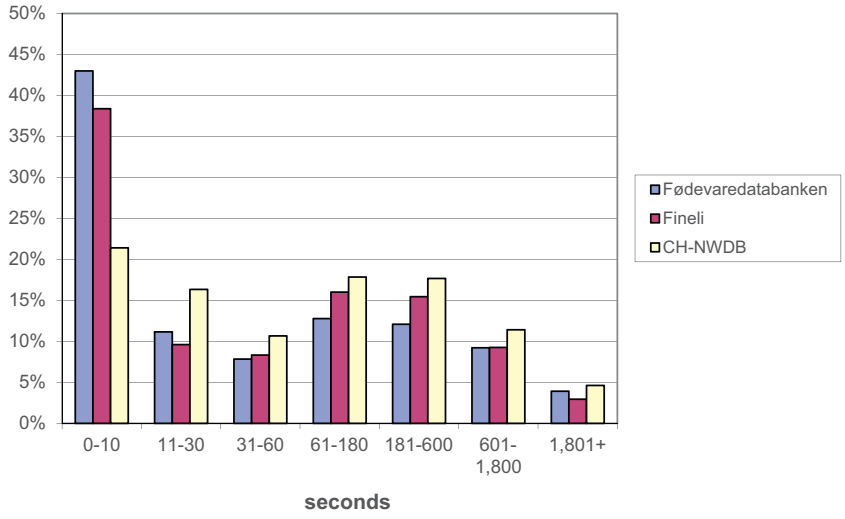

Fig. 4. Distribution of the visitors' time on site for the study websites, measuring period 1 May–30 November 2010. Top 100 web pages.

Table 6
Weighted average time on page when arriving from a referring site, grouped by main interest groups; weighted by the number of visits. The rows are ordered by the rank of visits (pooled rank based on visits for each of the study websites). Measuring period 1 May–30 November 2010. Top 100 website referrers.

Main interest group	Fødevaredatabanken $n = 20\,758$ visits	Fineli $n = 156\,611$ visits	CH-NWDB $n = 347$ visits
Other (including internal links)	5 min 19 s	4 min 50 s	11 min 55 s
Health and lifestyle	8 min 8 s	3 min 50 s	
News portal, newspapers, general portal	2 min 37 s	3 min 32 s	5 min 27 s
Wikipedia or other web-based encyclopaedia	1 min 52 s	1 min 44 s	–
Product promotion, webshop	3 min 37 s	2 min 12 s	30 min 10 s
Bodybuilders, weightlifters, fitness, training	1 min 32 s	2 min 43 s	–
Social media website (e.g. Facebook (Facebook, 2011))	3 min 21 s	3 min 15 s	8 min 27 s
Personal websites and blogs	9 min 34 s	2 min 28 s	9 min 0 s
Food education	14 min 35 s	10 min 50 s	–
General bulletin board	3 min 10 s	2 min 15 s	–
Home, cooking, gardening	8 min 28 s	4 min 15 s	–
Weight loss, dieting	2 min 28 s	4 min 33 s	–
Low carbohydrate diet	1 min 22 s	4 min 7 s	–
NGO food promotion (bread, fish, meat, etc.)	4 min 32 s	6 min 10 s	–
Pets, animal nutrition	2 min 51 s	2 min 20 s	–
Search engines	6 min 2 s	2 min 26 s	–
Peer support group (for diseases e.g. diabetes)	3 min 44 s	5 min 6 s	–
Food composition	5 min 45 s	4 min 55 s	–
Food counselling	5 min 49 s	4 min 12 s	–
Vegan, vegetarian	2 min 18 s	7 min 52 s	–
Babies, children, breastfeeding	2 min 46 s	2 min 48 s	–
Clinical nutrition	3 min 2 s	–	–
Science	5 min 42 s	3 min 44 s	–
Food safety	2 min 49 s	4 min 24 s	–
Unspecific diets	–	4 min 18 s	–

users interested in food counselling, science and food education (Fødevaredatabanken); and different diets or weight loss (Fineli).

Fig. 4 shows how much time visitors spent on the website. The distribution had two peaks, indicating two different groups of visitors: 'experimenters' who try the website and visit for less than 10 s and 'real' visitors who typically spent 1–10 min on the website. Moreover, quite a lot of 'real' visitors spent more than 10 min on the website (19.8–23.1% of 'real visitors').

3.6. What devices were used to visit our websites

The traditional workstation was the most common device used to visit the study websites (98.0–98.5% of visits). The amount of mobile users was in general very low on the study websites (1.5–2.0% of visits) and the number of pages per visit was also considerably lower compared with other types of devices. The visitors using tablet computers, however, visited about the same number of pages or even more pages (per visit) than users of workstation computers (Table 7). Moreover, the bounce rates seemed to be at about the same level. Currently, the total number of visitors using tablet computers was quite insignificant (only 0.1–0.2% of the visits).

Table 7
Pages per visit by different types of device. Measuring period 1 May–30 November 2010.

Operating system type	Fødevaredatabanken	Fineli	CH-NWDB
Mobile	4.1	4.9	9.1
Tablet	10.6	9.0	16.9
Workstation	9.0	9.9	13.9

4. Discussion

The results show that all the study websites had a significant number of visits particularly from their home country, though the range differed between countries, with the Finnish website receiving 100 times the number of visits compared to the Swiss website. The similar difference was found in the dissemination rate: the Fineli seems to reach a remarkable portion of the Finnish population (6.8%) compared with the others. This probably reflects the 'maturity' of the websites: Fineli is a long-established website with a high number of visits, which contrasts with the CH-NWDB website, which was launched 8 years later and until recently did not receive any active promotion in Switzerland. As the access to the Internet (Organisation for Economic Co-operation and Development (OECD), 2008) was more or less of the same order in all the study countries, the dissemination rate can be expected to be quite similar. Thus, there seems to be a large number of as yet unreached users in regard to Fødevaredatabanken and in particular to CH-NWDB. Therefore, facts or indicators as to why the number of visits differed in such a way are of particular interest to us. The results did not show any clear reason why CH-NWDB in particular gained so few visits. In general, the new visit rate is usually considered a measure of the promotional work that has been done for the site (Google Inc., 2009). On CH-NWDB, the new visit rate is only slightly lower than for the other websites. Moreover, the bounce rate on CH-NWDB was very low. With this combination, one would have expected more visits. The number of visits from search engines was higher for Fineli compared with the other websites. This indicates that Fineli was easier to find by occasional users and suggests that increased promotion could be useful for the other websites. In addition, the traffic of CH-NWDB consisted mostly of those users who accessed it directly, which stresses the need for promotion but also a greater need to take search engines into account in the website design.

As a general rule, access through search engines can be considered a good success indicator for the website: our results suggested that the more visitors who came to the website through search engines, the more popular was the website. This could be explained – at least partly – by the so-called snowball effect: the more traffic you have the more traffic you will gain. On the other hand, it could be stated that the more visitors a website had through direct traffic, the less visitors the website had in total. In general, the majority of the users seemed to know of the existence of the study websites as they commonly used the name of the website in the search. Use of the generic food composition search terms (e.g. 'food composition table') seemed to decrease with the increasing number of visitors. This could be explained perhaps by the fact the first visitors to a new food composition website were more 'food composition oriented professionals' and they tended to use the search terms that were 'appropriate' to the jargon of the field (i.e. 'professional' search terms). As the number of 'professionals' was rather limited, their number did not increase with the expanding visitor base. With the food name and the nutrient compound name, the situation was slightly better and their usage seemed to increase with increasing numbers of visitors. The results suggest that even when we have a considerable number of visitors, we should endeavour to know more about those potential users who have not yet found the website and still improve visibility in the search engines especially among those potential users who are not 'food composition oriented professionals'. This, however, also requires utilising other methods than measuring visitor statistics.

There were two types of visitors: those who just pop in and leave after a few seconds and those who spend several minutes upon arriving. The first group of visitors were probably looking for a different kind of content that what they found on the study websites, while the latter group could be considered as 'real' users. However, even when the visitors spent a reasonably long time on one of the study websites, it did not tell us whether they spent the time reading the content or trying to find it. On the other hand, the results from the most used content showed that most of the time was spent on the pages with food composition information – although the data do not give any indication as to whether the visitor found the content interesting or whether the content was so difficult that the visitor had to spend time trying to understand it. Then again, the visitors did not seem to use so much time navigating, even though the number of page visits on navigation-related pages was high. This gives reason to believe that the navigation and search facilities were not too difficult to use, though this topic needs further research. Moreover, the popularity of the personal food diary on Fineli suggests that a feature of this kind could be popular on other websites as well.

Visitor loyalty indicated that visitors found the content satisfactory enough that the website was worth visiting several times during the study period. About 15–20% of the visitors visited the website more than nine times during the relatively short measurement period. Visitor loyalty is an important issue to follow. As pointed by Wang, Pallister, and Foxall (2006), there are two extremely important visitor groups from the point of view of website design: first-time visitors, whose opinion and behaviour are important for understanding how to encourage a visitor to come back, and steady users, who would display the characteristics of visitors who return to the site regularly. Measuring visitor loyalty is key to finding and analysing these target groups. Visitor loyalty is often measured together with visitor recency (time between visits) (Croll & Power, 2009), although the quite short measurement period in the study (7 months) made its application unfeasible. The successful measurement of the multiple visits (i.e. unique visitors) is based on browser cookies and depends on whether a visitor allows these cookies or deletes them later, though this problem is not considered a major drawback (Kaushik, 2010).

Following traffic from the referring websites gave us more information about why users were visiting our website and what they were perhaps trying to accomplish. When a user entered our website from a website focused on a low carbohydrate diet, for instance, we did not know whether the visitor favoured the diet or not – but the diet very likely interested the visitor. Our results showed that a food composition website gained quite a broad spectrum of visitors and most of them could not be categorised as food or nutrition professionals. Many of the visitors seemed to be oriented towards health and lifestyle, physical exercise, or different diets. The websites were commonly used also for food education. Moreover, our websites seemed to be regular sources of information for such different groups as dog owners and people with diabetes. Visitors from several visitor groups had rather low bounce rates and they seemed to spend rather long time using the information provided to them indicating that the information was relevant to them. In addition, several these groups (e.g. food counselling, diets, dieting, food education, peer support groups for the diseases) could be concerned as target groups of which suggests that websites meet the expectations. Even though traffic from referring websites covered only a portion of all visits, it is still the only way to identify visitors. Referring websites are important in other aspects as well. With the help of GA, it is possible to access the website where our website was referred and read what the users of that website think of our website. In particular, following a conversation that takes place on a bulletin board that has a referral link may give valuable information about the 'overall reputation' of our website (Kaushik, 2010). Moreover, the referring websites are essential to search engines, as they are also taken into account when results of a search are shown. The more 'high ranked' websites that refer to your website, the better ranking your own website gets in the search results. The

benefit is obvious, as most search engine users use only the top results from the first result page (Croll & Power, 2009).

Even though use of mobile broadband (International Telecommunication Union, 2011) is quite common, only a small fraction of visitors accessed the websites by means of mobile devices. One of the reasons was that perhaps the current form of food composition data (large tables) was not suitable for devices with rather small screens, such as current smartphones. However, we did not find any difference in the number of viewed pages. Maybe food composition data can be offered to users via another kind of user interface or applications than the traditional browser interface. It is also possible to fine-tune the browser interface to be more suitable for mobile devices – this has not yet been attempted for the study websites. As the usage of mobile devices is estimated to grow further (Morgan Stanley, 2010), more attention should be paid to this potential user segment in future.

There is one methodological deficiency in GA. It is practically impossible to download any 'raw data' (not aggregated) of the visits. Even though there is a Data Export API for that purpose, it is possible to retrieve only very limited amounts of data (Google Inc., 2011). This means that the available data is mostly aggregated (daily, weekly) and thus it is not feasible to use it to study more detailed distributions. This is, however, not a serious insufficiency, as the goal is usually to monitor the website and to find issues that need improvement rather that to apply heavy statistical analysis. In addition, there are methodological problems to analysing visitor behaviour on a food composition website. When analysing transactional websites (such as web shops), it is common to create a so-called conversion funnel, where the ultimate goal (where the funnel points) is to get the visitor to make the purchase (Croll & Power, 2009). On an information-oriented website it is more difficult to define the conversion (i.e. what we would like the visitors do on our websites). In some cases, the number of downloaded white papers is used as a measure of conversion (Kaushik, 2010). So far, we have not found an appropriate conversion to be used on a food composition website. This needs profound knowledge of the users' needs and cannot be obtained only through website metrics (Croll & Power, 2009). Moreover, one of the overriding reasons for maintaining a food composition website is to provide information, and by some means, to promote healthier eating habits, it would be useful to measure whether we achieve something or not. It is perhaps overambitious to measure whether websites are able to contribute to solving problems in public health, but we should be able to measure whether we produce understandable information that is easy to find.

5. Conclusions

Our experiences show that GA is quite a versatile tool that gives very useful information in return for rather small effort and cost. However, GA is not the only tool available for doing web analytics. There are tools that function similarly to GA such as Yahoo! Web Analytics (Yahoo!, 2011), and tools like Piwik (Piwik, 2011), which can be installed on your own web server. Moreover, there are various commercial services or tools, some of which can be used alongside GA. Consequently, web analytics cannot be claimed to be difficult or expensive to implement or exploit. We think that monitoring website traffic and web analytics should be routine for every website – as well as on a food composite website, making user behaviour clearer so that developers can produce better websites for users.

Acknowledgement

We would like to thank Mr. Mark Phillips for linguistic revision.

References

Bhatnagar, A. (2009). Web analytics for business intelligence beyond hits and sessions. *Online*, *33*(6), 32–35.
Church, S. (2009). Food composition explained. EuroFIR Synthesis report No. 7. *Nutrition Bulletin*, *34*, 250–272.
Croll, A., & Power, C. (2009). *Complete web monitoring*. Sebastopol: O'Reilly Media Inc.
ETH Zurich. (2010). *Schweizer Nährwertdatenbank*. Retrieved from http://foodcase.ethz.ch
EuroFIR. (2010). *EuroFIR – Public homepage*. Retrieved from http://www.eurofir.eu/
Eurostat, (2011). Total population. Retrieved from http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&plugin=1&language=en&pcode=tps00001.
Facebook, (2011). Retrieved from http://www.facebook.com/.
Fang, W. (2007). *Using Google Analytics for improving library website content and design: A case study* (Electronic Version). Retrieved from http://works.bepress.com/wfang/1
Google Inc. (2009). *Google Analytics*. Retrieved from http://www.google.com/analytics/
Google Inc. (2011). *Data export API – Getting started*. Retrieved from http://code.google.com/apis/analytics/docs/gdata/gdataDeveloperGuide.html
Hasan, L., Morris, A., & Probets, S. (2009). Using Google Analytics to evaluate the usability of E-commerce sites. *Lecture Notes in Computer Science*, (5619), 697–706.
IBM. (2011). *IBM SPSS Statistics 19*. Retrieved from http://www.spss.com/software/statistics/
International Telecommunication Union. (2011). *Key global telecom indicators for the world telecommunication service sector*. Retrieved from http://www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom.html
Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research*, *28*, 407–432.
Kaushik, A. (2010). *Web Analytics 2.0*. Indianapolis: Wiley Publishing, Inc.
Morgan Stanley. (2010). *Internet trends*. Retrieved from http://www.morganstanley.com/institutional/techresearch/pdfs/Internet_Trends_041210.pdf
National Institute for Health and Welfare. Nutrition Unit. (2010). *Fineli – Finnish Food Composition Database. Release 11*. Retrieved from http://www.fineli.fi
Organisation for Economic Co-operation and Development (OECD). (2008). *The future of the Internet economy. A statistical profile*. Retrieved from http://www.oecd.org/dataoecd/44/56/40827598.pdf
Pakkala, H., Reinivuo, H., & Ovaskainen, M.-L. (2006). Food composition on the World Wide Web: A user-centred perspective. *Journal of Food Composition and Analysis*, *19*(2–3), 231–240.
Peterson, E. T. (2004). *Web analytics demystified. A marketer's guide to understanding how your web site affects your business*. Celilo Group Media and CafePress.
Peterson, E. T. (2006). *The big book of key performance indicators*. Portland/Boston/Madrid: Web Analytics Demystified.
Piwik. (2011). *Piwik. Open source web analytics*. Retrieved from http://piwik.org/
Plaza, B. (2009). Monitoring web traffic source effectiveness with Google Analytics: An experiment with time series. *Aslib Proceedings*, *61*(5), 474–482.
Plaza, B. (2009b). *Using Google Analytics for measuring inlinks effectiveness*.
Plaza, B. (2011). Google Analytics for measuring website performance. *Tourism Management*, *32*(3), 477–481.
Technical University of Denmark. (2010). *Fødevaredatabanken. Version 7.01*. Retrieved from http://www.foodcomp.dk/
W3Techs. (2010). *Usage of traffic analysis tools for websites*. Retrieved from http://w3techs.com/technologies/overview/traffic_analysis/all
Wang, H.-C., Pallister, J. G., & Foxall, G. R. (2006). Innovativeness and involvement as determinants of website loyalty III. Theoretical and managerial contributions. *Technovation*, *26*, 1374–1383.
Wikimedia Foundation Inc. (2011). *Wikipedia. The free encyclopedia*. Retrieved from http://www.wikipedia.org/
Williamson, C. (2005). The different uses of food composition databases. Synthesis report No. 2 (Electronic Version). EuroFIR Synthesis report from http://www.eurofir.net/sites/default/files/EuroFIR%20synthesis%20reports/Synthesis%20Report%202_The%20different%20uses%20of%20food%20composition%20databases.pdf.
Yahoo!. (2011). *Yahoo! Web Analytics*. Retrieved from http://web.analytics.yahoo.com/

Heikki Pakkala is a data analyst in the Nutrition Unit of the National Institute for Health and Welfare, Finland. He holds a master degree in Ecology from University of Oulu, Finland. Currently he works as a software engineer, especially for the Finnish National food composition database. Moreover, he has worked in the European Food Information resource (EuroFIR) Network of Excellence and is currently a director in EuroFIR AISBL (www.eurofir.org). Prior to the current position he has over 10 years expertise from private IT-sector. His articles have appeared in Journal of Food Composition and Analysis, European Journal of Clinical Nutrition, Apetite.

Karl Presser, Dipl. Informatik-Ing. ETH works as a Ph.D. student at the Department of Computer Science at the ETH Zurich. He worked 4 years in an SME as database designer, software architect and product manager. As part of his Ph.D. project he developed a food composition management software (FoodCASE), which is now available for use by European food composition data managers, thus enabling a standardised working tool across Europe, and which is part of the FP7 EuroFIR Nexus project. The main research objectives in his Ph.D. project

are to define, measure, and manage data quality aspects in scientific database systems.

Tue Christensen holds a masters degree in food science from Royal Veterinarian and Agricultural University of Denmark and has a position as senior adviser at the National Food Institute at Technical University of Denmark. He leads the nutrition data processing group, which tasks involve compiling and handling food composition data, data capture for dietary analysis, reporting of dietary intake and risk analysis. He has worked with systems dealing with food and food composition data since 1984. He has published articles in Journal of Food Composition and Analysis, Public Health Nutrition, European Journal of Clinical Nutrition, etc.