# Wikipedia Tools for Google Spreadsheets

Thomas Steiner
Google Germany GmbH
ABC Str. 19, 20354 Hamburg, Germany
tomac@google.com

## **ABSTRACT**

In this paper, we introduce the Wikipedia Tools for Google Spreadsheets. Google Spreadsheets is part of a free, Webbased software office suite offered by Google within its Google Docs service. It allows users to create and edit spreadsheets online, while collaborating with other users in realtime. Wikipedia is a free-access, free-content Internet encyclopedia, whose content and data is available, among other means, through an API. With the Wikipedia Tools for Google Spreadsheets, we have created a toolkit that facilitates working with Wikipedia data from within a spreadsheet context. We make these tools available as open-source on GitHub, released under the permissive Apache 2.0 license.

# **Categories and Subject Descriptors**

H.3.5 [Online Information Services]: Web-based services

# **Keywords**

Wikipedia, Wikidata, Google Spreadsheets, Google Sheets

# 1. INTRODUCTION

In the world of Computer Science, spreadsheet applications serve for the organization, analysis, and storage of data in tabular form. Spreadsheets are the computerized simulation of paper accounting worksheets, and operate on data represented as cells of an array, organized in rows and columns. Cells can contain numeric or textual data, or the results of formulas that automatically calculate and display a value based on the contents of other cells. With the Wikipedia Tools for Google Spreadsheets, we introduce a toolkit of such formulas, tailored to the universe of Wikipedia, that enables a wide range of potential use cases starting from marketing, to search engine optimization, to business analysis. Especially through the chaining of formulas, the true power and ease of spreadsheet applications can be unleashed.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.

WWW'16 Companion, April 11-15, 2016, Montréal, Québec, Canada.

ACM 978-1-4503-4144-8/16/04. http://dx.doi.org/10.1145/2872518.2891112

# 1.1 Wikipedia and Wikidata

Wikipedia's content and data is available through the Wikipedia API (https://{language}.wikipedia.org/w/api.php), where {language} represents one of the currently 291 supported Wikipedia languages,² for example, en for English, de for German, or zu for Zulu. Wikidata is a collaboratively edited knowledge base and intended to provide a common source of structured data which can be used by projects such as Wikipedia. Its content and data is available through the Wikidata API (https://www.wikidata.org/w/api.php). Both the Wikipedia and the Wikidata APIs' data is available as XML or JSON, among other formats. Wikipedia pageviews data, i.e., the number of times within a given period of time that a given Wikipedia article has been viewed can be obtained using the Pageviews API (https://wikimedia.org/api/rest\_v1/?doc). The data is available in JSON format.

# 1.2 Google Spreadsheets and Apps Scripts

Google Spreadsheets can be extended with custom functions (or formulas) using Google Apps Scripts<sup>3</sup> that are written in standard JavaScript.<sup>4</sup> To illustrate this, a trivial function is defined in Listing 1 that can then be used from within a spreadsheet as outlined in Listing 2. Custom functions can access external resources on the Web by fetching URLs with the UrlFetchApp, one of the scripting services available in Google Apps Script. Fetched data can either be in XML or JSON format and parsed with convenience functions.

```
function DOUBLE(input) {
  return input * 2;
}
```

Listing 1: Custom Google Sheets function called DOUBLE.

=DOUBLE (A1)

Listing 2: Usage of the custom DOUBLE function from Listing 1 in a cell with the value of cell A1 as a parameter.

### 2. LIST OF DEVELOPED FUNCTIONS

In our Wikipedia Tools for Google Spreadsheets, we provide eleven functions that—in traditional spreadsheets style—follow an all-uppercase naming convention and start with

 $^2{\rm List}$  of Wikipedias: https://meta.wikimedia.org/wiki/List\_of\_Wikipedias

<sup>3</sup>Google Apps Script: https://developers.google.com/apps-script/

<sup>4</sup>Custom functions in Google Sheets: https://developers.google.com/apps-script/guides/sheets/functions

<sup>&</sup>lt;sup>1</sup> Wikipedia Tools for Google Spreadsheets: https://github.com/tomayac/wikipedia-tools-for-google-spreadsheets

a WIKI prefix. These functions are wrappers around the particular Wikipedia or Wikidata API calls, or the Pageviews API respectively. Figure 1 shows exemplary output for the English Wikipedia article https://en.wikipedia.org/wiki/Berlin and the English Wikipedia category https://en.wikipedia.org/wiki/Category:Berlin. The functions are listed below.

**WIKITRANSLATE** Returns Wikipedia translations (language links) for a Wikipedia article.

**WIKISYNONYMS** Returns Wikipedia synonyms (redirects) for a Wikipedia article.

**WIKIEXPAND** Returns Wikipedia translations (language links) and synonyms (redirects) for a Wikipedia article.

**WIKICATEGORYMEMBERS** Returns Wikipedia category members for a Wikipedia category.

**WIKISUBCATEGORIES** Returns Wikipedia subcategories for a Wikipedia category.

**WIKIINBOUNDLINKS** Returns Wikipedia inbound links for a Wikipedia article.

**WIKIOUTBOUNDLINKS** Returns Wikipedia outbound links for a Wikipedia article.

**WIKIMUTUALLINKS** Returns Wikipedia mutual links, i.e, the intersection of inbound and outbound links for a Wikipedia article.

**WIKIGEOCOORDINATES** Returns Wikipedia geocoordinates for a Wikipedia article.

**WIKIDATAFACTS** Returns Wikidata facts for a Wikipedia article.

**WIKIPAGEVIEWS** Returns Wikipedia pageviews statistics for a Wikipedia article.

**WIKIPAGEEDITS** Returns Wikipedia pageedits statistics for a Wikipedia article.

Most functions directly wrap native API calls, with three exceptions: (i) the functionality of the WIKISYNONYMS and the WIKITRANSLATE functions is combined in the WIKIEXPAND function, both the WIKITRANSLATE and the WIKIEXPAND function accept an optional target languages parameter that allows for limiting the output to just a subset of all available Wikipedia languages; (ii) the function WIKIMUTUALLINKS is the intersection of the two functions wikiinboundlinks and WIKIOUTBOUNDLINKS; and (iii) the function WIKIDATAFACTS provides a list of claims [11] (or facts), enriched with entity and property labels for improved readability, limited to single-value objects, and simplified using an adapted version of Maxime Lathuilière's simplifyClaims function<sup>5</sup> from his Wikidata SDK [6]. This allows us to return two columns in RDF [2] terms "predicate" and "object" pairs—with one unique object, for example, the predicate ISO 3166-2 code with the object DE-BE, and deliberately discarding multivalue claims, for example, predicate head of government with objects Michael Müller and Klaus Wowereit, among many others. While in the concrete example the ordering is clear (temporal), this is not true in the general case, for example, with predicate instance of. As a result, in WIKIDATAFACTS, we prefer indisputability of claims over their completeness. Listing 3 exemplarily shows the complete implementation of the WIKISYNONYMS function.

```
* Returns Wikipedia synonyms
 * @param {string} article The Wikipedia article
 * @return {Array<string>} The list of synonyms
function WIKISYNONYMS(article) {
  'use strict';
  if (!article)
    return '';
  var results = [];
  trv {
    var language = article.split(/:(.+)?/)[0];
    var title = article.split(/:(.+)?/)[1];
    if (!title) {
      return '';
    title = title.replace(/\s/g, '_');
    var url = 'https://' + language +
        '.wikipedia.org/w/api.php' +
        '?action=query' +
        '&blnamespace=0' +
        '&list=backlinks' +
        '&blfilterredir=redirects' +
        '&bllimit=max' +
        '&format=xml' +
        '&bltitle=' +
        encodeURIComponent(title);
    var xml = UrlFetchApp.fetch(url)
        .getContentText();
    var document = XmlService.parse(xml);
    var entries = document.getRootElement()
        .getChild('query').getChild('backlinks')
        .getChildren('bl');
    for (var i = 0; i < entries.length; i++) {
      var text = entries[i].getAttribute('title')
          .getValue();
      results[i] = text;
  } catch (e) {
    // no-op
  return results.length > 0 ? results : ":
```

Listing 3: Implementation of WIKISYNONYMS.

#### 3. USAGE SCENARIOS

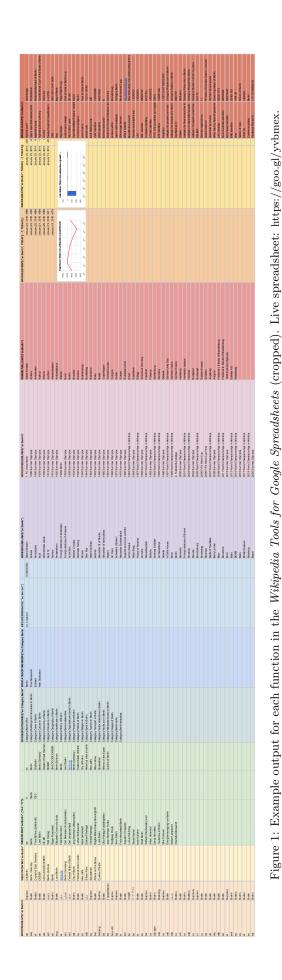
We have tested the Wikipedia Tools for Google Spreadsheets with different usage scenarios in mind. These include, but are not limited to, the ones listed in the following.

# 3.1 Usage Scenario I: Ordered Category Panel

Wikipedia holds an enormous amount of categories, for example, visitor attractions in Montreal.<sup>6</sup> Category members obtained through a call of WIKICATEGORYMEMBERS are listed in alphabetical order, however, if we additionally request pageviews data for each category member through a series of WIKIPAGEVIEWS calls and then sort by pageviews in descending order, we get a representative list of top-10 visitor attractions—enriched with photos retrieved through calls of WIKIDATAFACTS filtered on "image"—as shown in Figure 2. A similar feature (based on non-disclosed metrics) in form

 $<sup>^5</sup> Wikidata SDK$  simplify Claims function: https://github.com/maxlath/wikidata-sdk#simplify-claims-results

 $<sup>^6 \</sup>mbox{Visitor}$  attractions in Montreal: https://en.wikipedia.org/wiki/Category:Visitor\_attractions\_in\_Montreal



of an image carousel can be seen in Google's Knowledge Graph [10] Web search results pages when searching for "visitor attractions in montreal" (demo https://goo.gl/Ugt0je).

# 3.2 Usage Scenario II: Search Ads

Search advertisers can greatly profit from the information that is contained in Wikipedia and Wikidata. For example, if we imagine a hotel booking site, it may be desirable to advertise based on points of interest (POIs) and create advertisements automatically featuring known facts of such POIs. Figure 3 shows an example where skyscrapers listed in the category skyscrapers over 350 meter<sup>7</sup> are first obtained via WIKICATEGORYMEMBERS and then checked for their "height" fact via WIKIDATAFACTS, which is then used in two templates to create ads. Search keywords are generated by calling WIKISYNONYMS and combined with terms like "hotel".

# 3.3 Usage Scenario III: Marketing Campaigns

On January 13, 2016, Google Maps added Street View imagery for the model railway *Miniatur Wunderland*. Taking global Wikipedia pageviews as a popularity indicator, we can examine if the marketing campaign has had any impact on the attraction, assuming that more pageviews translate to increased visitor interest. Therefore, we first obtain the *Miniatur Wunderland* article in all available languages via WIKITRANSLATE and then retrieve pageviews via WIKITRAGEVIEWS. Figure 4 shows indeed an international uptake of pageviews starting January 13 after an earlier linear curve progression (except for the German article, which had a peak on January 8, a long weekend after a public holiday).

## 4. RELATED WORK

In his book Google Apps Script for Beginners [4], Gabet gives an introduction to extending Google Spreadsheets with custom functions. A similar introduction is given in Ferreira's Google Apps Script: Web Application Development Essentials [3]. In [5], Han et al. describe their approach RDF123 to translate spreadsheets data to RDF, the inverse of what we do in WIKIDATAFACTS. Olsen and Moser show in [8] how Web APIs can be taught with spreadsheets. The process of calling Web APIs via spreadsheets is further described in [9]. Further, in [1], Abramson et al. describe how they enabled spreadsheets to have "super-computing" powers through parallelized custom functions. An open-source toolkit for mining Wikipedia—not bound to spreadsheets, but designed for general use with the Java programming language—is described by Milne et al. in [7].

# 5. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced the Wikipedia Tools for Google Spreadsheets. First, we have introduced the data sources Wikipedia and Wikidata and their different APIs. Second, we have shown how Google Spreadsheets can be extended through custom functions that can then be used from within a cell context as if they were native functions. In the following, we have listed the implemented functions, and explained where they extend the functionality of the underly-

 $<sup>^7 \</sup>rm Skyscrapers$  over 350 meter: https://en.wikipedia.org/wiki/Category:Skyscrapers\_over\_350\_meters

<sup>&</sup>lt;sup>8</sup>Miniatur Wunderland on Google Street View https://www.google.com/maps/about/behind-the-scenes/streetview/treks/miniatur-wunderland/

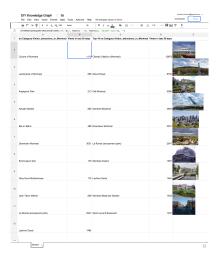


Figure 2: Usage scenario I: Wikipedia Tools for Google Spreadsheets used to create an ordered category panel based on Wikipedia category memberships and accumulated Wikipedia pageviews for popularity ranking (here: the top-10 visitor attractions in Montreal). Live spreadsheet: https://goo.gl/Njvt1T.

		Artist	10 B Z + A	A ⊞ ⊞ ± ± 1	÷ ∞ ■   □ ▼ Σ	
fic.	*ARREST GRALLA (LONES (VEXT SYNEWING (*ve	"AC\$2)A"	hotel"))			
	A		c			,
1	en:Category: Skyserapers over 350 meter	Holett (ex	0			
2	1 Park Avenue		Potencies Travers	International Commerce Centre	Shanghei World Financial Center	Taipel 101
2	One World Trade Center		1452	1454	1450	4500
÷	111 West 57th Street		Hotel near Petrones Towers		Hotel near Shanghai World Financial Cont	
i	23 Marina		Enjoy the view from 452m above		Enjoy the view from 492m above	Enjoy the view from 509m above
÷	Those World Trade Carder		Sionscrapers over 350 meters	Skysprapers over 350 meters	Skyscrapers peer 350 meters	Skyscrapers over 350 meters
,	Abrai Al Balt	+601			www.exemple.org/hotelishanghai-world-finan-	
÷	Abu Phahi Plana		and the party of t			annie garage and
è	Akhmet Tower		Petronas Towers Hotel	International Commerce Corne III	Shanetel World Financial Center Hotel	Taipel 101 Hotel
70	Al Harma Tower		Raiax 452m above the ground		Relax 482m above the ground	Relax 500m above the ground
11	Al Dudy Enginement Tracer		Get a hotel with a perfect view	Get a hotel with a perfect view	Get a hotel with a perfect view	Get a hotel with a perfect view
12	Al Baihi Tower				www.exemple.org/hotelishanghai-world-finan-	
13	Almas Tinear		and the party of t			
14	Acers Towns		petronas tower hotel	union source chase 7 hotel	shanohai world financial centre hotel	taipei 101 tower hotel
15	Bank of America Tower (Manhattan)		petronas towers hotel	international commerce center heng		101 tower hotel
*	Bank of China Tower (Hong Kong)		petronas twin towers hotel	international commerce center hotel		tainel financial center hotel
g	Bacneng Shensang Global Financial Center		petronas tower 1 hotel	kowloon station phase 7 hotel	world's largest bottle opener hotel	101 tower hotel
18	Birea Treer		petronas tower 2 hotel	ritz-ceriton (hong kang) hotel	sharohei wfc hotel	taipel financial centre hotel
90	Bin Managa Twin Towers		petronis twin towers hotel	international commerce centre hong		taipsi international francial centre ho
20	Buri Al Alam		petronas twin towers hotel	international commerce centre hong		tainei international financial center to
21	Buri Khalifa		husta lumpur hein towers hotel		shinghili huingi) (hrong phonosin hotel	talkai financial cantra hotal
22	Buri Mubarak al-Kabir		bein towers (malaysia) botel		shanohei financial center hotel	taibei financial center hotel
23	Busian Lotte Town Tower				shanghei hills world financial center hotel	talkel international financial centre has
24	Capital Market Authority Headquarters				shanghai world finance center hotel	tabel international francial center to
25	Central Market Project					taibei yilingyi hotel
26	Central Park Tower					talbel quoii limono zhongvin hotel
27	Central Plaza (Hong Kong)					taibei 101 hotel
20	Changsha IF8 Tower T1					taineh 101 hotel
20	Cheroply Greenland Tower					tainel tower hotel
30	China Resources Centre Block A					St 101 hotel
21	China Resources Headquarters					62.8101 hotel
32	China Zun					taibili vilinovi hotel
33	Chongoing International Trade and Commer					Cal-pel Hing-I hotel
34	CITIC Plaza					tal-oak fileno-fi hotel
25	CTF Finance Centre	+530				tapel 101 hotel
20	Dalian Plan Center					taipei 101 building hotel
37	Dallan Greenland Center					
30	Dalian International Trade Center					
29	DAMAC Residence					
-	Diseased Toward Institute					

Figure 3: Usage scenario II: Wikipedia Tools for Google Spreadsheets used to create textual search ads based on Wikidata facts (here: skyscraper heights) and Wikipedia synonyms as keywords combined with the term "hotel". Live spreadsheet: https://goo.gl/np1Is8.

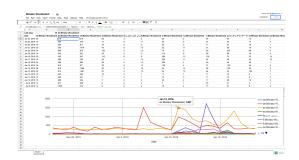


Figure 4: Usage scenario III: Wikipedia Tools for Google Spreadsheets used to evaluate the impact of a marketing campaign (here: model railway Miniatur Wunderland being featured on Google Street View since January 13, 2016). Live spreadsheet: https://goo.gl/q1yhuV.

ing wrapped API functions. We have then focused on three different usage scenarios that illustrate how to work with the *Wikipedia Tools for Google Spreadsheets* and finally have provided an overlook on related work in the area.

Future work will focus on adding more functions as need be and potentially making the functions more parameterizable. In the current iteration, we have favored simplicity and ease of use over customizability, essentially making the most common use case the only option. Possibly, in upcoming releases, we will add an advanced mode that allows experienced users to fine-tune the functions' results, for example, to implicitly include bot traffic in WIKIPAGEVIEWS that we have currently excluded on purpose.

Concluding, we were positively surprised by the increased productivity and short turnaround time enabled by the Wikipedia Tools for Google Spreadsheets for the rapid prototyping of ideas, especially in combination with the fill-down and fill-right features in spreadsheets and the charting capabilities. We look forward to making the tools even more powerful and hope to attract collaborators for the open source project available on GitHub at https://github.com/tomayac/wikipedia-tools-for-google-spreadsheets. As a positive side effect, the tools can even help improve Wikipedia and Wikidata when authors add missing data, for example, we added an image to one of the visitor attractions of Montreal, as this fact was initially missing in Wikidata (and thus in Figure 2).

## 6. REFERENCES

- D. Abramson, L. Kotler, D. Mather, and P. Roe. ActiveSheets: Super-Computing with Spreadsheets. In U. Seattle, editor, Proceedings of the High Performance Computing Symposium - HPC 2001, pages 110-115, San Diego, USA, 2001.
- [2] R. Cyganiak, D. Wood, and M. Lanthaler. RDF 1.1 Concepts and Abstract Syntax. Recommendation, W3C, Feb. 2014.
- [3] J. Ferreira. Google Apps Script: Web Application Development Essentials. O'Reilly Media, 2014.
- [4] S. Gabet. Google Apps Script for Beginners. Packt Publishing, 2014.
- [5] L. Han, T. Finin, C. Parr, J. Sachs, and A. Joshi. RDF123: From Spreadsheets to RDF. In *The Semantic Web – ISWC 2008*, volume 5318 of *LNCS*, pages 451–466. Springer, 2008.
- [6] M. Lathuilière. Wikidata SDK, 2016. https://github.com/maxlath/wikidata-sdk (2016-02-08).
- [7] D. Milne and I. H. Witten. An Open-Source Toolkit for Mining Wikipedia. Artificial Intelligence, 194:222–239, Jan. 2013.
- [8] T. Olsen and K. Moser. Teaching Web APIs in Introductory and Programming Classes: Why and How. Paper 16, SIGED: IAIM Conference, Feb. 2013.
- [9] K. Patel, S. Prish, S. Sadhu, L. Bizek, and X. Pan. Spreadsheet Functions to Call REST API Sources, May 15 2014. US Patent App. 13/672,704.
- [10] A. Singhal. "Introducing the Knowledge Graph: things, not strings", Official Google Blog, May 2012. http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.
- [11] D. Vrandečić and M. Krötzsch. Wikidata: A Free Collaborative Knowledgebase. Commun. ACM, 57(10):78–85, Sept. 2014.