

別紙：目次

【付録 1】 検索回数による二分化基準設定条件	1
1-A 「求める文献を的確に発見できないユーザー」	1
2-B 「求める文献を的確に発見できるユーザー」	2
【付録 2】 演算子使用による二分化基準設定条件	3
2-A 「求める文献を的確に発見できないユーザー」	3
2-B 「求める文献を的確に発見できるユーザー」	4
【付録 3】 比較視点におけるデータ抽出・集計方法	6
検索語	6
端末	8
滞在時間	8
所属	8
演算子を使った検索	9
検索方法	10
参照元	11
Referral	11
【付録 4】 GoogleQueryExplorer の使い方	11
使用法	12
GQE の設定例	15
参考ページ	16

【付録 1】 検索回数による二分化基準設定条件

調査は、2014年9月1日から11月10日にかけて行われた。従って、Google Analyticsのインターフェースや設定もその時点の環境を利用した。

※青字は入力・選択項目

1-A 「求める文献を的確に発見できないユーザー」

- ◆ CiNii Articles を 1 回検索し、検索結果一覧を眺めて終了するユーザー 1 セッションあたり、サイト内ページを 2 ページ閲覧し、かつセッションがトップページ (/) で始まり、次に簡易検索・詳細検索結果一覧 (/search?) か著者検索結果一覧 (/author?) か全文検索結果一覧 (/fulltext?) のいずれかを通過してサイトから出て行った行動を Google Analytics の「セグメント」機能で設定する。

具体的な設定条件は次の通りである。

- 設定条件

- 【条件】

- フィルタ：セッションを含める

- ページビュー数=2

- 【シーケンス】

- フィルタ：セッションを含める

- シーケンスの開始：すべての接点

- ステップ 1：ランディングページ、完全一致、/

- (ステップ 1) 直後のステップ：ステップ 2

- ステップ 2：「離脱ページ、先頭が一致、/search?」 OR 「離脱ページ、先頭が一致、/author?」 OR 「離脱ページ、先頭が一致、/fulltext?」

- ◆ CiNii Articles を 1 回検索し、1 回論文詳細画面を見て終了するユーザー 1 セッションあたり、サイト内ページを 3 ページ閲覧し、かつセッションがトップページ (/) で始まり、次に簡易検索・詳細検索結果一覧 (/search?) か著者検索結果一覧 (/author?) か全文検索結果一覧 (/fulltext?) のいずれ

かを通過し、次に論文詳細画面を閲覧してサイトから出て行った行動を Google Analytics の「セグメント」機能で設定する。

具体的な設定条件は次の通りである。

- 設定条件

- 【条件】

- フィルタ：セッションを含める

- ページビュー数=3

- 【シーケンス】

- フィルタ：セッションを含める

- シーケンスの開始：すべての接点

- ステップ 1：ランディングページ、完全一致、/

- (ステップ 1) 直後のステップ：ステップ 2

- ステップ 2：「ページ、先頭が一致、/search?」 OR 「ページ、先頭が一

- 致、/author?」 OR 「ページ、先頭が一致、/fulltext?」

- (ステップ 2) 直後のステップ：ステップ 3

- ステップ 3：離脱ページ、先頭が一致、/naid/

2-B 「求める文献を的確に発見できるユーザー」

- ◆ CiNii Articles を 1 回以上検索し、その後も論文情報を閲覧し続けたユーザー

1 セッションあたり、サイト内ページを 4 ページ以上閲覧し、かつセッションがトップページ (/) で始まり、次に簡易検索・詳細検索結果一覧

(/search?) か著者検索結果一覧 (/author?) か全文検索結果一覧

(/fulltext?) のいずれかを通過した行動を Google Analytics の「セグメント」機能で設定する。

具体的な設定条件は次の通りである。

- 設定条件

- 【条件】

- フィルタ：セッションを含める

ページビュー数>=4

【シーケンス】

フィルタ：セッションを含める

シーケンスの開始：すべての接点

ステップ 1：ランディング ページ、完全一致、/

(ステップ 1) 直後のステップ：ステップ 2

ステップ 2：「ページ、先頭が一致、/search?」 OR 「ページ、先頭が一致、/author?」 OR 「ページ、先頭が一致、/fulltext?」

【付録 2】演算子使用による二分化基準設定条件

調査は、2014 年 9 月 1 日から 11 月 10 日にかけて行われた。従って、Google Analytics のインターフェースや設定もその時点の環境を利用した。

※青字は入力・選択項目

2-A 「求める文献を的確に発見できないユーザー」

◆ 演算子を使わずに検索したユーザー

1 セッションあたり、AND 検索以外の演算子を使わずに検索を実行した行動を Google Analytics の「セグメント」機能で設定する。

具体的な設定条件は次の通りである。

● 設定条件

【条件】

フィルタ：セッションを除外する

検索キーワード、正規表現に一致、".+"

OR 検索キーワード、正規表現に一致、"/.+"/

OR 検索キーワード、正規表現に一致、".+¥*"

OR 検索キーワード、正規表現に一致、¥s(OR|¥)|¥s

OR 検索キーワード、正規表現に一致、¥sNOT¥s

OR 検索キーワード、正規表現に一致、¥s¥-.+"

OR 検索キーワード、正規表現に一致、
 $\$(.+ \$s(\$s|\& \$s|OR \$s|¥| \$s|NOT \$s|-).+ \$)$

フィルタ 2: セッションを含める
ページ、先頭が一致、/search?

OR ページ、先頭が一致、/author?

OR ページ、先頭が一致、/fulltext?

2-B 「求める文献を的確に発見できるユーザー」

◆ AND 検索以外の演算子を使って検索できるユーザー

1 セッションあたり、AND 検索以外の演算子を使って検索を実行した行動を Google Analytics の「セグメント」機能で設定する。
具体的な設定条件は次の通りである。

● 設定条件

【条件】

フィルタ: セッションを含める

検索キーワード、正規表現に一致、".+"

OR 検索キーワード、正規表現に一致、/./

OR 検索キーワード、正規表現に一致、.+¥*

OR 検索キーワード、正規表現に一致、 $\$(OR|¥)| \s

OR 検索キーワード、正規表現に一致、 $\$(NOT \$s)$

OR 検索キーワード、正規表現に一致、 $\$(¥)-.+$

OR 検索キーワード、正規表現に一致、

$\$(.+ \$s(\$s|\& \$s|OR \$s|¥| \$s|NOT \$s|-).+ \$)$

フィルタ 2: セッションを含める

ページ、先頭が一致、/search?

OR ページ、先頭が一致、/author?

OR ページ、先頭が一致、/fulltext?

- 補足

設定条件の各正規表現の意味については次の表を参照のこと。

正規表現	説明
"."+	フレーズ検索を表している。 CiNii Articles はアルファベット、数字などの1バイト文字にのみ対応。
././	完全一致を表している。
.+.*	前方一致を表している。 CiNii Articles はアルファベット、数字などの1バイト文字にのみ対応。
¥s(OR ¥)¥s	「OR」と「 」を使ったOR検索を表している。 CiNii Articles は半角・全角文字、使用可。
¥sNOT¥s	「NOT」を使った場合のNOT検索を表している。 CiNii Articles は半角・全角文字、使用可。
¥s¥-.+	「-」を使った場合のNOT検索を表している。
¥(.+¥s(¥s &¥s OR¥s ¥ ¥s NOT¥s -).+¥)	丸かっこ「()」を使って、論理演算の優先度を指定した場合を表している。 丸かっこ内には必ず演算子が含まれる。 CiNii Articles は半角・全角文字、使用可。

Google Analytics で使用可能な正規表現についてはヘルプページを参照した。

Google Inc.. “正規表現の概要 - アナリティクス ヘルプ”. Google アナリティクス. (オンライン),
<https://support.google.com/analytics/answer/1034324?hl=ja>, (参照 2014-12-09)

また、Perl の正規表現に近い。

Google Inc. . “Core Reporting API – Reference Guide – Google Analytics — Google Developers” . Google アナリティクス. (オンライン) , <https://developers.google.com/analytics/devguides/reporting/core/v3/reference>, (参照 2014-12-05)

【付録 3】 比較視点におけるデータ抽出・集計方法

【付録 1】 及び【付録 2】 では二分化基準のユーザーを Google Analytics のセグメント機能を使ってセグメントごとに絞り込む設定条件を説明した。ここでは、比較視点での傾向の比較を行うためのデータ抽出・集計方法を説明する。なお、「求める文献を的確に見えないユーザー」である「CiNii Articles を 1 回検索し、検索結果一覧を眺めて終了するユーザー」と「CiNii Articles を 1 回検索し、1 回論文詳細画面を見て終了するユーザー」は別々のセグメントとして設定している。そのため、セグメントごとにデータを抽出・集計し、後で合算したものを該当ユーザーとしている。

検索語

Google Analytics のレポートメニュー「行動」>「サイトコンテンツ」>「すべてのページ」で、プライマリディメンションに「ページ」、セカンダリディメンションに「検索キーワード」を設定。アドバンスフィルタで「ページ」に「`^*/search?q`」を含むデータを簡易検索データとして、「`^*/fulltext?q`」を含むデータを全文検索データとして絞りこみ、セグメントごとにファイルに出力した。

ファイルに出力したデータを Microsoft Access (以下 Access) に取込む。検索行動を伴わない検索結果一覧の 2 ページ以降を閲覧したデータを除外し、1 ページ目を閲覧したデータのみを抽出するために、「ページ」に引数「`&start`」を含むデータを除外した。また、簡易検索データから詳細検索データを除外するため、詳細検索で使用する引数を含むデータを除外した。下表は、詳細検索で使用する引数の一覧である。

日本語名称	引数
タイトル	title
著者名	author
著者所属	affiliation
刊行物名	journal
ISSN	issn
巻	volume
号	issue
ページ	page
出版者	publisher
参考文献	references
出版年 ～から	year_from
出版年 ～まで	year_to

これらのデータに基づいて、以下の調査1、2を行った。

◆ 調査1

抽出したデータを Microsoft Excel（以下 Excel）を使用して集計する。検索語の分割には And 検索がスペースで区切られている性質を利用した。

「検索キーワード」フィールドにある単語を Excel の区切り分割機能でスペースごとに1語ずつに分割する。分割された検索語の単語数を、関数「COUNTIF(範囲, "*"")+COUNT(範囲)」を利用してカウントする。記号も1語にカウントされる場合があるが、そのままカウントしている。単語全体の長さは、関数「SUMPRODUCT(LEN(範囲))」でカウントした。単語数の平均値は[単語数×ページビュー数の合計]÷[単語数の合計]、検索語全体の長さの平均値は[単語の長さ×ページビュー数の合計]÷[単語数の合計]、検索語一語当たりの長さの平均値は[単語の長さ×ページビュー数の合計]÷[単語数×ページビュー数の合計]で求めた。

◆ 調査2

調査1の簡易検索のデータを利用して集計。単語数ごとにページビュー

数を集計し、さらにその構成比を算出した。

また、調査1、2とは違ったやり方で調査3を行った。

◆ 調査3

Google Analytics のレポートメニュー「行動」>「サイト内検索」>「サイト内キーワード」で、プライマリディメンションに「検索キーワード」、セカンダリディメンションに「再検索キーワード」を設定。アドバンスフィルタで「検索キーワード」に該当検索語を含むデータを絞りこみ、検索語ごとにファイルに出力した。

抽出したデータを Microsoft Excel（以下 Excel）を使用して集計する。検索語の分割には And 検索がスペースで区切られている性質を利用した。

「検索キーワード」「再検索キーワード」フィールドにある検索語を Excel の区切り分割機能でスペースごとにそれぞれ1語ずつに分割する。分割された単語の語数を、関数「COUNTIF(範囲, ” *”)+COUNT(範囲)」を利用してカウントする。記号も1語にカウントされる場合があるが、そのままカウントしている。1回目検索の「検索キーワード」と2回目検索の「再検索キーワード」の語数を比較することで、1回目から2回目にかけての語数の増減を判定する。また、1回目検索と2回目検索の1番目、2番目の単語について、それぞれ同じであるか異なっているかを判定した。

端末

Google Analytics のレポートメニュー「ユーザー」>「モバイル」>「サマリー」でプライマリディメンションに「デバイスカテゴリ」を設定して、ファイルに出力。さらにその構成比をセグメントごとに算出した。

滞在時間

Google Analytics のレポートメニュー「ユーザー」>「サマリー」>「平均セッション時間」に表示されるデータを参照した。

所属

Google Analytics のレポートメニュー「ユーザー」>「ユーザーの環境」>「ネットワーク」でプライマリディメンションに「サービスプロバイダ」、セカンダリディメンション

に「ネットワークドメイン」を設定して、セグメントごとにファイルに出力する。

ファイルに出力したデータを Access に取込み、「学術機関」とみなす条件を、ネットワークドメインが ac.jp で終わるもの、及びサービスプロバイダの名称から学術機関と判断できるもの（名称に Universty、College、Institute を含むもの）を「学術機関」とし、データを抽出。全体のセッション数の合計から「学術機関」のセッション数の合計を引いたものを「その他」とし、「学術機関」と「その他」の構成比をセグメントごとに算出した。

演算子を使った検索

Google Analytics のレポートメニュー「行動」>「サイトコンテンツ」>「すべてのページ」でプライマリディメンションに「ページ」を設定して、セグメントごとにファイルに出力する。

ファイルに出力したデータを Access に取込み、次の一覧の抽出条件と判断基準で、「検索キーワード」フィールドでの演算子の使用を特定し、ページビュー数の合計を算出。各演算子の合計とすべての演算子の合計について、全体に対する割合を求めた。

	抽出条件	CiNii Articles ルール	判断基準
フレーズ検索	Like “ ” * ” ” ”	アルファベット、 数字などの1パイ ト文字にのみ対 応	2バイト文字でもカウ ント 前後セットになってい ない場合はカウントし ない 全体を囲んでいる場合 のみカウント
完全一致	Like “*/ */*”		前後セットになってい ない場合はカウントし ない
前方一致	Like “*[*]”	アルファベット、 数字などの1パイ ト文字にのみ対 応	2バイト文字でもカウ ント
AND 検索	Like “* *” Or Like “*[*]”	半角・全角文字、 使用可	
OR 検索	Like “*OR *” Or Like “* [] *”	半角・全角文字、 使用可	大文字のみカウント
NOT 検索	Like “* NOT *” Or Like “* -*”	半角・全角文字、 使用可	大文字のみカウント ハイフンで囲まれてい る、ハイフンが連続して いる場合はカウントし ない
丸かっこ検索	Like “*(* NOT *) *” Or Like “*(* [-] *) *” Or Like “*(* OR *) *” Or Like “*(* [] *) *” Or Like “*(* *) *” Or Like “*(* [&] *) *”	半角・全角文字、 使用可	前後セットになってい ない場合、カッコ内に演 算子が含まれない場合 はカウントしない

検索方法

Google Analytics のレポートメニュー「行動」>「サイトコンテンツ」>「すべてのページ」でプライマリディメンションに「ページ」を設定する。アドバンスフィルタで「ペ

ージ」について、「`^¥/search¥?`」で簡易検索及び詳細検索データ、「`^¥/fulltext¥?q`」で全文検索データ、「`^/author¥?q`」で著者検索データを絞り込んで抽出し、セグメントごとにファイルに出力する。

ファイル出力したデータを Access に取込む。検索行動を伴わない検索結果一覧の 2 ページ以降を閲覧したデータを除外し、1 ページ目を閲覧したデータのみを抽出するために、「ページ」に引数「`&start`」を含むデータを除外した。さらに、簡易検索及び詳細検索データについては、「ページ」に「`^/search¥?q`」を含むデータのうち詳細検索の引数を含まないものを簡易検索データとして抽出し、差分を詳細検索データとした。詳細検索の引数については「**検索語**」の項を参照のこと。

参照元

Google Analytics のレポートメニュー「集客」>「チャンネル」でプライマリディメンションに「Default Channel Grouping」を設定して、ファイルに出力。セグメントごとに構成比を算出した。

Referral

Google Analytics のレポートメニュー「集客」>「チャンネル」>「Referral」プライマリディメンションに「参照元」を設定して、セグメントごとにファイルに出力。フィールド「ドメイン名」に「`ac.jp`」を含むものを「大学等」、ドメイン名から学術情報に関するサイトと判断できるものを「一般学術系サイト」、残りを「その他」と分類し、セグメントごとに構成比を算出した。

【付録 4】 GoogleQueryExplorer の使い方

調査は、2014 年 9 月 1 日から 11 月 10 日にかけて行われた。従って、Google Analytics のインターフェースや設定もその時点の環境を利用した。

Google Query Explorer（以後、GQE）とは Google Analytics におけるデータ出力をサポートするツール機能で、Google Analytics で使用されているデータを任意の区分（dimensions）と指標（metrics）で出力することができるものである。Google Analytics のインターフェースで出力されるリストはもちろんのこと、独自に考案した区分と指標を設定することで、Google Analytics のインターフェースでは出力できないリストも利用することができるようになる。Google Analytics にはデータ出力用の API や各プログラム言語で

使用できるライブラリーなど様々なツール/アイテムがあるが、自身で別のツールを用意せずに Web インターフェースだけでデータを出力できる点で GQE は簡易なツールと言える。

使用法

GQE は Google Analytics が利用できる Google アカウントを設定し、集計したい区分と指標、対象期間等を設定してデータを出力させるだけなので特別な使用方法はないのだが、Google Analytics との指標の関連付けなどが少し難解なので、対応する項目と簡単な設定方法を以下に紹介する。

まずは次ページに GQE の Web インターフェースを挙げる。(2014. 12. 9 現在)

Google Analytics Demos & Tools

Click here to authorize →

Query Explorer

Overview

Sometimes you just need to explore. This tool lets you play with queries to get data from your Google Analytics views (profiles). You can also use the client libraries to build your own tools.

Account:

Property:

View (Profile):

* ids =

dimensions =

* metrics =

segment =

filters =

sort =

* start-date =

* end-date =

start-index =

max-results =

ids
(Required)
The namespaced view (profile) ID of the view (profile) from which to request data. Use the view (profile) selector above to

1174 is your view (profile) ID.

[Read the reference for ids](#)

Google Analytics technology used in this tool
The Query Explorer uses the [Core Reporting API](#) to query for data and create the reports. It uses the [Management API](#) to get a list of accounts and segments the current user can access, and it uses the [Metadata API](#) to populate the items in the dimension and metrics dropdowns.

ユーザー
アカウント
&
レポート部



データ設定部



データ出力部



◆ ユーザーアカウント & レポート部

Google Analytics で使用しているユーザーアカウントと、レポート（分析対象の Web サイト）を設定する。ユーザーアカウントの認証を行えば、利用可能なレポートをプルダウンメニューで選択することが可能。

◆ データ設定部

出力したいデータの区分(dimensions)や指標(metrics)、その他の項目について設定することができる。各項目については、以下を参照。

- ・ dimensions : 出力データの主区分。プライマリディメンションとされることもあり、Google Analytics のデータ表示（リスト）では行の項目となるもの。複数選択可。複数選択する場合は、セカンダリディメンション等の設定と解釈される。

- ・ metrics : 出力データの指標。Google Analytics のデータ表示（リスト）では列の項目となるもの。複数選択可。選択された分だけ出力される項目が増える。

- ・ segment : 出力データに対するセグメント設定。デフォルトのセグメント設定の他、認証させているユーザーアカウントで Google Analytics 上で作成したセグメントについても利用可能。複数選択不可。

- ・ filters : 出力データに対する絞り込み設定。特定のキーワードを含むものや、あるキーワードから始まるもの、等任意の設定で出力データを絞り込むことができる。正規表現も利用可能。

- ・ sort : 出力データの並び替え設定。dimensions や metrics で指定した項目を指定することで並び替えられたデータで出力可能となる。逆順も指定可。

- ・ start-date/end-date : 分析対象となるログデータの期間設定。

- ・ start-index : 出力するデータの最初の順位指定。GQE では最大 10000 行までしかデータを出力できないため、10001 位以降のデータを取得するためには、本項目で 10001 と指定し

て出力することが必要となる。(なお、Google Analytics では最大 5000 行しか取得できない)

・max-results: 出力行数の指定。任意に指定可能だが、GQE では一回に最大 10000 行までしかデータを取得できないため、実質上限は 10000 となる。

・データ設定のコツ: 各項目のプルダウンメニューで指標等を選択する際、全て英語のラベルなので理解しにくいですが、Google Analytics 側の設定を日本語から英語にして、まず Google Analytics で出したい表の dimensions や metrics の英語名を調べ、それから GQE のプルダウンメニューの中から探すことで対応ラベルを見つけることができる。

◆ データ出力部

必要な設定を行った上で「Get Data」ボタンを押下するとデータが表形式で出力される。表示されたデータは TSV 形式 (タブ区切りテキスト) で出力可能。

GQE の設定例

本調査で使用した比較項目のいくつかについて、GQE でデータ出力した際の設定例を挙げる。

◆ 「所属」

(レポートメニュー「ユーザー」>「ユーザーの環境」>「ネットワーク」)

プライマリディメンションに「サービスプロバイダ」、セカンダリディメンションに「ネットワークドメイン」を設定)

```
dimensions: ga:networkLocation, ga:networkDomain
```

```
metrics:
```

```
ga:sessions, ga:percentNewSessions, ga:newUsers, ga:bounceRate, ga:pageviewsPerSession, ga:avgSessionDuration
```

```
sort: -ga:sessions
```

```
start-date: 2013-09-01
```

```
end-date: 2014-08-31
```

◆ 「参照元」

(レポートメニュー「集客」>「チャンネル」

プライマリディメンションに「Default Channel Grouping」を設定)

```
dimensions: ga:channelGrouping
```

```
metrics:
```

```
ga:sessions, ga:percentNewSessions, ga:newUsers, ga:bounceRate, ga:pageviewsPerSession, ga:avgSessionDuration
```

```
sort: -ga:sessions
```

```
start-date: 2013-09-01
```

```
end-date: 2014-08-31
```

◆ 「端末」

(レポートメニュー「ユーザー」>「モバイル」>「サマリー」

プライマリディメンションに「デバイスカテゴリ」を設定)

```
dimensions: ga:deviceCategory
```

```
metrics:
```

```
ga:sessions, ga:percentNewSessions, ga:newUsers, ga:bounceRate, ga:pageviewsPerSession, ga:avgSessionDuration
```

```
sort: -ga:sessions
```

```
start-date: 2013-09-01
```

```
end-date: 2014-08-31
```

参考ページ

・セルティス電子工房. “Google Analytics Query Explorer 2 | セルティスラボ” セルティスラボ. (オンライン),

<http://celtislab.net/archives/20130717/google-analytics-query-explorer-2/>, (参照 2014-12-10)

少し前のフォーマットだが、参考になる。

・Google Inc.. “Core Reporting API - Reference Guide - Google Analytics — Google Developers”. Google Analytics Core Reporting API - Reference Guide. (オンライン),

<https://developers.google.com/analytics/devguides/reporting/core/v3/reference>,

(参照 2014-12-10)

API の説明だが、dimensions や metrics についての説明は共通のようだ。