

**Étude des pratiques et besoins numériques
des utilisateurs aux Archives de l'Etat, à la
Bibliothèque royale et au CegeSoma
(2015-2017)**

Les données « Web Analytics »

Anne Chardonens

Octobre 2017





| | |
|--|------------|
| <u>INTRODUCTION</u> | 4 |
| <u>CORPUS</u> | 5 |
| <u>METHODOLOGIE</u> | 7 |
| CHOIX DE L’OUTIL | 7 |
| PIWIK | 9 |
| GESTION DES URLS | 9 |
| ACCES AUX DONNEES BRUTES | 11 |
| ANALYSE QUANTITATIVE DE L’AUDIENCE EN LIGNE | 14 |
| TROIS NIVEAUX D’ANALYSE..... | 15 |
| METHODE DES “GOALS” | 15 |
| SOLUTIONS SUR MESURE | 18 |
| VIGILANCE | 19 |
| <u>PARTIE I : SITES WEB INSTITUTIONNELS</u> | 20 |
| INTRODUCTION | 21 |
| 1.1. VUE COMPARATIVE : SITES WEB INSTITUTIONNELS | 22 |
| VISITES ET VISITEURS..... | 22 |
| REPARTITION DES VISITES..... | 23 |
| 1.2. ANALYSES SPECIFIQUES : SITES WEB INSTITUTIONNELS | 33 |
| 1.2.1. ANALYSE SPECIFIQUE : WWW.ARCH.BE | 33 |
| 1.2.2. ANALYSE SPECIFIQUE : WWW.KBR.BE | 60 |
| 1.3.3 ANALYSE SPECIFIQUE : WWW.CEGESOMA.BE..... | 84 |
| <u>PARTIE II : SITES DES CATALOGUES EN LIGNE, SITES D’ACCES AUX COLLECTIONS NUMERISEES ET SITES DE VALORISATION DES COLLECTIONS</u> | 106 |
| INTRODUCTION | 107 |
| 2.1. VUE COMPARATIVE : SITES DES CATALOGUES EN LIGNE ET SITES D’ACCES AUX COLLECTIONS NUMERISEES | 109 |
| 2.1.1. VISITES ET VISITEURS | 109 |
| 2.1.2. REPARTITION DES VISITES..... | 110 |



| | |
|--|------------|
| 2.2. VUE COMPARATIVE : SITES DE VALORISATION DES COLLECTIONS..... | 127 |
| 2.3. ANALYSES SPECIFIQUES : SITES DES CATALOGUES EN LIGNE ET SITES D'ACCES AUX COLLECTIONS NUMERISEES | 139 |
| 2.3.1. ANALYSE SPECIFIQUE : SEARCH | 139 |
| 2.3.2 ANALYSE SPECIFIQUE : OPAC KBR..... | 158 |
| 2.3.3. ANALYSE SPECIFIQUE : BELGICA..... | 172 |
| 2.3.4. ANALYSE SPECIFIQUE : PALLAS | 179 |
| 2.3.5. ANALYSE SPECIFIQUE : WARPRESS..... | 190 |
| <u>CONCLUSION</u> | 197 |



Introduction

Ce document vise à présenter le volet “Web Analytics” du projet MADDLAIN, autrement dit, l’analyse des traces d’usage collectées sur les différents sites web de trois établissements scientifiques fédéraux. Cette analyse de données quantitatives s’inscrit dans une enquête plus large, qui a pour objectif de mieux connaître et de mieux comprendre les utilisateurs de ces institutions, en vue de pouvoir offrir un meilleur accès à l’offre numérique.

Deux doctorants en Sciences et Technologies de l’Information et de la Communication (ULB)¹ se sont succédés pour travailler sur cette problématique : Raphaël Hubain (mars à décembre 2015) et Anne Chardonens (janvier à novembre 2016).

Ce rapport est composé de deux parties principales : la première repose sur une analyse globale des comportements des utilisateurs sur les différents sites web des institutions, tandis que la seconde se concentre sur les catalogues en ligne. Chaque partie comporte une description de la méthodologie utilisée et le détail des résultats obtenus.

Enfin, il faut garder à l’esprit que si les données de fréquentation des sites web permettent de répondre à la question du “quoi”, elles ne suffisent pas pour répondre à celle du “pourquoi” (les intentions de l’utilisateur). Cette dernière nécessite de recourir à une approche plus qualitative. La dernière phase du projet MADDLAIN permettra donc de combiner ces deux types de données, afin d’avoir une compréhension plus fine du comportement des utilisateurs.

¹ Rattaché au MaSTIC ULB : <http://mastic.ulb.ac.be>

Corpus

Les données ont été collectées dès l'été 2015² par le biais de Piwik³, alternative open source à Google Analytics, qui permet notamment de rester propriétaire des données. L'intégration de Piwik, la gestion des cookies, la collecte et le stockage des données ont été réalisés par Martin Vanbrabant et Gerald Haesendonck, deux chercheurs de Imec⁴.

L'extraction et le traitement des données ont été effectués à l'aide de Piwik, MySQL, Excel, OpenRefine et Pandas (bibliothèque Python). Les données collectées proviennent de 11 sites web :

| INSTITUTION | NOM | DESSCRIPTIF | URL |
|---------------------------------|----------------------------|---|---|
| CegeSoma | CGS-www | Site général | http://www.cegesoma.be |
| CegeSoma | Pallas | Catalogue en ligne | http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm |
| CegeSoma | Warpress | Accès à la presse de guerre numérisée (belge) | http://warpress.cegesoma.be |
| CegeSoma | Journal of Belgian History | Site de la Revue belge d'Histoire contemporaine | http://www.journalbelgianhistory.be |
| Archives de l'État en Belgique | AGR-www | Site général | http://www.arch.be |
| Archives de l'État en Belgique | Search | Catalogue en ligne | http://search.arch.be |
| Archives de l'État en Belgique | Expo Congo | Exposition virtuelle | http://www.expocongo.be |
| Archives de l'État en Belgique | Goaaal ! | Exposition virtuelle | http://goaaal.arch.be |
| Bibliothèque royale de Belgique | KBR-www | Site général ⁵ | http://www.kbr.be |

² Dans le cas des AGR, certains sites étaient déjà *trackés* par Piwik depuis un certain temps.

³ <http://www.piwik.org>

⁴ Centre de recherche et d'innovation né de la fusion entre iMinds et imec : www2.imec.be

⁵ L'analyse effectuée dans le cadre du projet MADDLAIN concerne l'ancien site web de la KBR, qui a été remplacé par une nouvelle version le 15 septembre 2016.



| | | | |
|---------------------------------|---------|------------------------|---|
| Bibliothèque royale de Belgique | Opac | Catalogue en ligne | http://opac.kbr.be |
| Bibliothèque royale de Belgique | Belgica | Bibliothèque numérique | http://belgica.kbr.be |

Il est important de souligner que les données présentées au sein de ce rapport incluent toutes les visites réalisées sur ces sites, y compris celles qui ont été effectuées par des membres du personnel et par des personnes utilisant un ordinateur mis à disposition au sein de l'institution.⁶

⁶ En théorie, il est possible d'exclure certaines adresses IP via Piwik. Cependant, au-delà des postes de travail fixes auxquels peut être associé une adresse IP, subsisterait la question des adresses IP dynamiques et des connexions effectuées au sein de l'institution via des ordinateurs, tablettes ou téléphones personnels.

Méthodologie

La collecte des données de fréquentation des sites web s'est organisée en plusieurs étapes : choix de l'outil le plus approprié, capture des traces de visite pour les pages web avec URLs uniques, réflexion et solution pour pouvoir également enregistrer les visites sur les pages web comportant des URLs ambiguës (plusieurs pages web comportent la même adresse) et enfin, récupération des données brutes concernant l'ensemble de ces visites.

Onze sites web ont été analysés dans le cadre du projet MADDLAIN, provenant des trois établissements scientifiques étudiés : le CegeSoma, les Archives de l'État et la Bibliothèque royale de Belgique. Il s'agit de sites de contenu général, d'expositions virtuelles et d'interfaces permettant aux utilisateurs d'effectuer des recherches dans les catalogues en ligne des institutions afin d'accéder aux collections numérisées ou non.

Choix de l'outil

Lorsque qu'il est question du comportement des utilisateurs sur un site web, l'outil d'analyse le plus populaire est Google Analytics. Pour permettre à Google Analytics d'enregistrer ce qui se déroule sur un site web, il suffit d'effectuer une modification sur chaque page du site. Des consignes simples permettent d'insérer un code Javascript. À chaque visite sur cette page, des données seront envoyées à Google Analytics à partir du navigateur web du visiteur. Ces données de visites reprennent notamment l'adresse de la page (URL), le titre de la page, l'adresse de la page précédente (sur laquelle un lien faisait référence à cette page), l'adresse IP de l'utilisateur (qui, dans un souci d'anonymat, est partiellement masquée) et enfin toutes les données concernant le navigateur, le système d'exploitation, et la taille de l'écran de l'appareil avec lequel se connecte l'utilisateur.

Ces données sont traitées par Google Analytics, qui les utilisera afin de déduire de nouvelles informations : Comment le visiteur est-il arrivé sur le site web (moteurs de recherche, adresse directe, redirection depuis un autre site web) ? Quelle page du site a été utilisée comme "page d'entrée" ? Quelles sont les pages qui ont été successivement visitées par l'utilisateur et quel a été le temps passé sur chacune de ces pages ? Quelle page du site s'est avérée être la "page de sortie" de l'utilisateur ? Quelle est la localisation de l'utilisateur ?



Ces informations sont fournies par Google Analytics sous formes de rapports contenant des données statistiques. Concrètement, cela signifie que Google Analytics offre un aperçu du temps passé en moyenne par les utilisateurs sur une page, une représentation géographique de la provenance des visiteurs pour une période donnée, ou encore le “bounce rate” par page, c’est-à-dire le pourcentage d’internautes ayant quitté le site immédiatement après n’avoir vu qu’une seule page du site. Étant donné que la quantité d’informations à traiter ne cesse d’augmenter, Google Analytics résume ces informations et les présente sous forme de rapports peu détaillés (processus d’agrégation) ou ne génère certains rapports que pour une partie des visiteurs (processus d’échantillonnage).

L’ensemble de l’environnement Google Analytics fonctionne à l’aide des serveurs de Google. Cela signifie que les utilisateurs du service ne sont pas obligés de posséder leur propre infrastructure. Ce mode de fonctionnement comporte cependant d’importantes implications pour l’utilisateur final : non seulement aucune responsabilité personnelle, mais pas non plus de contrôle ni de garantie de fonctionnalité cohérente à long terme.

Par exemple, Google a récemment procédé à une transition de “Google Analytics” vers “Universal Analytics”, ce qui a notamment entraîné un ajustement du fichier javascript utilisé⁷ ainsi que l’émergence de “Google Tag Manager”, qui permet d’ajouter de nouveaux outils d’analyse sur les pages web. Or, le code source Google Analytics est détenu par Google. Il s’agit donc de déterminer si les rapports fournis par Google Analytics ou Universal Analytics sont suffisants pour l’analyse qui est prévue. Dans le cadre du projet MADDLAIN, ces rapports n’étaient pas jugés suffisants et les options payantes proposées par Universal Analytics à ceux qui souhaitent obtenir une exportation des données brutes n’étaient pas encore disponibles en 2015, au début du projet. L’option d’une solution alternative a donc dû être étudiée.

⁷ ga.js est ainsi devenu analytics.js, afin de permettre la mise en place de fonctionnalités additionnelles, cf. <https://developers.google.com/analytics/devguides/collection/upgrade/reference/gajs-analyticsjs>

Piwik

L'équipe du projet a opté pour Piwik, une alternative populaire et "open source" à Google Analytics⁸. Piwik inclut des fonctionnalités similaires à celles proposées par Google Analytics, avec le même principe d'insertion de code Javascript sur les pages web, ce qui permet ensuite d'enregistrer les données de fréquentation sur le site et de générer des rapports. La principale différence réside dans le fait que la personne utilisant Piwik garde à tout moment accès aux données brutes, qui peuvent être utilisées afin de réaliser une analyse détaillée. Elle peut également décider si des données brutes doivent être effacées ou non, ce qui permet d'avoir un contrôle direct sur ce qui doit être inclus ou non dans les rapports générés par Piwik. Enfin, vu que Piwik est un logiciel "open source", les utilisateurs ne sont pas condamnés à se soumettre à une logique d'enfermement propriétaire ("vendor lock-in"). Concrètement, Piwik a été installé sur un serveur dévolu au projet au printemps 2015⁹. Tant le code Piwik que la base de données sont donc restés entre les mains des chercheurs impliqués dans le projet.

Gestion des URLs

L'adresse des pages Web visitées (l'URL) fait partie des données les plus cruciales en ce qui concerne les "données utilisateur". En effet, la plupart des rapports se servent de cette adresse comme d'un index auquel d'autres données sont liées. La majorité des sites web de contenu diffusent l'information en utilisant le principe d'URLs uniques, c'est-à-dire qu'à chaque page web correspond une URL propre. Par exemple, l'URL <http://www.arch.be/index.php?l=fr> correspond à la page d'accueil du site des Archives de l'État dans sa version francophone. Bien que les informations affichées puissent changer au fil du temps, chaque visiteur qui se rendra sur cette page à un temps T verra exactement le même contenu. À chaque clic du visiteur (par exemple sur le lien "Chercheur" dans la rubrique "Votre profil"), une nouvelle URL unique s'affichera, permettant de suivre le parcours du visiteur sur une nouvelle page : <http://www.arch.be/index.php?l=fr&m=chercheur>.

⁸ "Open source" signifie que le code du programme informatique est accessible à tous et peut être lu, modifié ou redistribué.

⁹ Ce serveur fait partie de l'infrastructure de Imec (anciennement iMinds), un centre de recherche partenaire du projet Maddlain.

D'autres types d'applications web permettent, grâce à une programmation spécifique, de modifier le contenu à afficher en fonction de l'interaction du visiteur avec le site web. Il existe deux types principaux¹⁰ : Une page web sur laquelle les changements de contenus activent un script qui est contenu dans le code HTML de la page et une page web sur laquelle le contenu est modifié après l'intervention du serveur web.

Dans le premier cas de figure, il est certain que l'URL correspond clairement au contenu affiché ; dans le second cas, il est possible qu'il n'y ait pas d'adéquation entre l'URL et le contenu affiché. Cela peut rendre l'analyse plus ardue. Un exemple de cette seconde configuration est le moteur de recherche du CegeSoma, Pallas, que l'on retrouve à cette adresse :

<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htm>.

Arrivé sur cette page, le visiteur peut : entrer un mot-clé, effectuer une recherche avancée, déterminer le type de contenus à rechercher et lancer la recherche.

Après avoir lancé sa recherche, il verra une nouvelle page s'afficher. En revanche, l'adresse (URL) n'a pas changé. Ce type de pages web s'avère problématique si l'on souhaite analyser l'activité qui se déroule sur le site. Ainsi, en l'absence d'intervention spécifique pour contourner ce problème d'URL ambiguës, les rapports concernant ces pages ne fourniront que très peu d'informations, peu importe qu'il s'agisse Google Analytics ou Piwik. En résumé, toutes ces "pages" seront regroupées sous la même URL et l'information fournie à ce sujet sera peu significative.

La solution consiste à enregistrer les actions des visiteurs. C'est-à-dire que pour chaque action donnant lieu à l'affichage d'un nouveau contenu, des données supplémentaires sont inscrites sur le serveur pour enregistrer en détail ce qui se passe au cours des visites. Ces données peuvent être enregistrées à l'aide de codes Javascript supplémentaires et de variables personnalisées ("custom variables").

Si cette mise en œuvre a lieu sur un site préexistant, sans que sa structure puisse être modifiée, un important travail préparatoire est nécessaire. Il s'agit d'observer de façon détaillée les différentes options présentes sur chaque page et d'analyser la source des

¹⁰ En outre, la combinaison des deux types est possible.

potentielles variations de ces pages. Ensuite, le code Javascript rédigé pour capturer ces actions doit être testé scrupuleusement. Cette tâche, qui exige beaucoup de discipline et de patience, doit être documenté de façon exhaustive en vue des analyses ultérieures. Elle a néanmoins été réalisées dans le cadre du projet MADDLAIN pour cinq sites web¹¹ sur les 12 analysés.

Accès aux données brutes

La base de données Piwik a été stockée sur un serveur interne au projet. Le choix du système de gestion de bases de données s'est porté sur *MariaDB*, une alternative open source à MySQL, utilisant une syntaxe de commandes SQL similaires. Piwik écrit dans cette base de données en utilisant des commandes SQL. Ces mêmes commandes SQL permettent ensuite de lire l'ensemble des tables de la base de données. Les données brutes sont donc situées dans des tables dont le nom commence par `piwik_log`.

Voici un exemple de ce qui peut être extrait de ces tables, dans le cadre d'une visite de Pallas (le moteur de recherche du CegeSoma). La requête suivante est destinée à faire apparaître l'horodatage, l'URL et les custom variables (informations additionnelles sur les actions du visiteur) générées pour la venue d'un visiteur donné au cours d'un intervalle de temps donné :

¹¹ Le catalogue numérique de la Bibliothèque royale <http://opac.kbr.be> et sa bibliothèque numérique, Belgica <http://belgica.kbr.be/> (qui n'est plus utilisée depuis la refonte du site kbr.be, achevée en septembre 2016) ; le catalogue numérique des Archives search.arch.be ; Pallas, le catalogue numérique du Cegesoma, <http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm> et Warpress, le site de valorisation de journaux numérisés <http://warpress.cegesoma.be>.

Fig. 3

REQUÊTE SQL

```
SELECT piwik_log_link_visit_action.server_time, piwik_log_action.name as url,
piwik_log_link_visit_action.custom_var_k1, piwik_log_link_visit_action.custom_var_v1,
piwik_log_link_visit_action.custom_var_k2, piwik_log_link_visit_action.custom_var_v2,
piwik_log_link_visit_action.custom_var_k3, piwik_log_link_visit_action.custom_var_v3,
piwik_log_link_visit_action.custom_var_k4, piwik_log_link_visit_action.custom_var_v4,
piwik_log_link_visit_action.custom_var_k5, piwik_log_link_visit_action.custom_var_v5
FROM piwik_log_link_visit_action
JOIN piwik_log_action
ON piwik_log_link_visit_action.idaction_url = piwik_log_action.idaction
JOIN piwik_log_visit
ON piwik_log_link_visit_action.idvisit = piwik_log_visit.idvisit
WHERE piwik_log_link_visit_action.idsite = 8 /* Cgs-pallas */
AND piwik_log_link_visit_action.server_time > "2017-02-10 14:40:00"
AND piwik_log_link_visit_action.server_time < "2017-02-10 14:50:00"
AND HEX(piwik_log_visit.idvisitor) = "aa47a8f0322a5207"
ORDER BY piwik_log_link_visit_action.server_time;
```

Voici les informations qui apparaissent suite à l'envoi de cette requête :

RÉSULTATS DE LA REQUÊTE SQL

Structure :

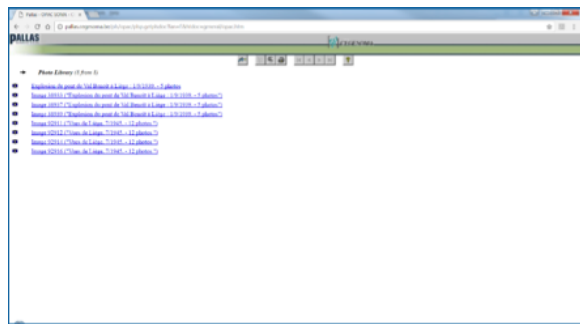

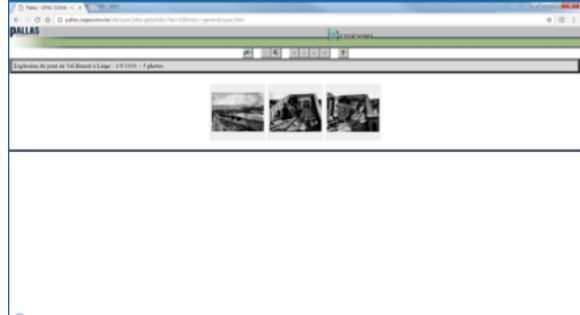

```
server_time    url    custom_var_k1 custom_var_v1 custom_var_k2 custom_var_v2
custom_var_k3 custom_var_v3 custom_var_k4 custom_var_v4 custom_var_k5
custom_var_v5
```

```
2017-02-10 14:44:23
pallas.cegesoma.be/pls/opac/plsp.getplsd
oc?lan=E&htdoc=general/opac.htm NULL
NULL NULL NULL NULL
NULL NULL NULL NULL
```



```
2017-02-10 14:44:38
pallas.cegesoma.be/pls/opac/plsp.getplsd
oc?lan=E&htdoc=general/opac.htm
FormData text=pont de Val
Benoît&action=search&Seop=5&in=_CA
NULL NULL NULL NULL NULL
NULL NULL NULL
```



| | |
|---|--|
| <p>2017-02-10 14:44:47 pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm page descriptionList link /pls/opac/opac.search?lan=E&seop=5&sele=3&sepa=1&doty=_CA&sest=pont de val benoit&chna=&senu=78234&rqdb=1&dbnu=1 linkText Explosion du pont de Val Benoit à Liège : 1/9/1939. - 5 photos NULL NULL NULL NULL</p> |  |
| <p>2017-02-10 14:44:51 pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm page detailInfo link window.location.href='/pls/opac/opac.search?lan=E&seop=5&sele=50&sepa=1&doty=_CA&sest=pont+de+val+benoit&chna=&senu=78234&rqdb=1&dbnu=1'; linkType ViewButton NULL NULL NULL NULL</p> |  |
| <p>2017-02-10 14:44:55 pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm page detailInfo link http://pallas.cegesoma.be/pls/opac/opac.search?lan=E&seop=5&sele=51&sepa=1&doty=_CA&sest=pont+de+val+benoit&chna=&senu=38933&rqdb=1&dbnu=1 linkText image http://pallas.cegesoma.be/pallas/servlet/git?vn=38933 NULL NULL</p> |  |
| <p>2017-02-10 14:45:00 pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm navigation Next image (2 of 3) NULL NULL</p> |  |

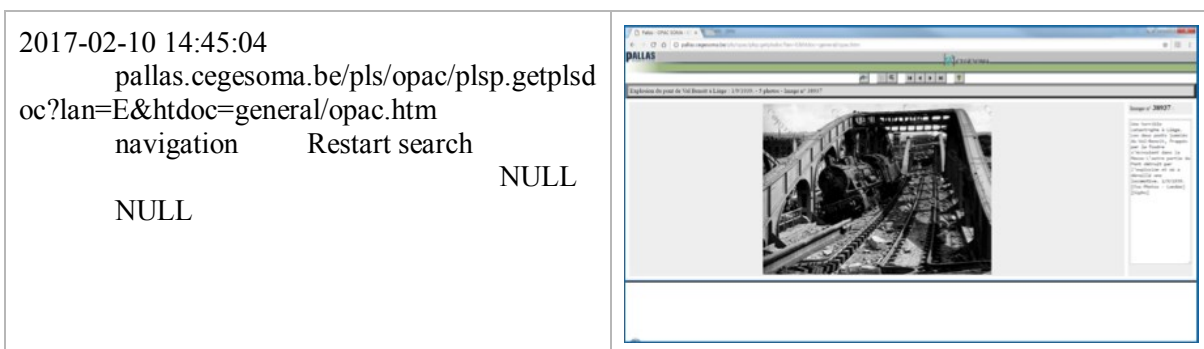


Fig. 3

Il s'agit des données brutes pour un utilisateur ayant effectué les actions suivantes :

- Commence sa visite sur Pallas
(<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm>)
, entre comme termes de recherches "pont de Val Benoît", sélectionne "Tout (combinaison)" comme critère de recherche et "Photothèque" comme sous-domaine de recherche.
- Lance la recherche
- Sur la page suivante qui s'affiche : clique sur le premier lien
- Sur la page suivante qui s'affiche : clique sur le bouton "Visualiser document(s)"
- Sur la page suivante qui s'affiche : clique sur la première miniature
- Sur la page suivante qui s'affiche : clique sur l'icône "page suivante"
- Sur la page suivante qui s'affiche : clique sur l'icône "page précédente".

Ces différents procédés permettent donc d'enregistrer le comportement des utilisateurs avec suffisamment de détails pour obtenir une image complète incluant notamment l'utilisation des collections numérisées mises à disposition par l'institution. Si la solution reste laborieuse, elle représente cependant une option viable dans les situations où ce type de mesures n'a pas été pensé en amont.

Analyse quantitative de l'audience en ligne

Après que les traces d'usage laissées par les utilisateurs ont été collectées, elles peuvent être analysées. Bien qu'elles ne permettent pas de connaître les intentions des utilisateurs, elles sont utiles pour savoir ce qui se passe sur les sites web des institutions, et cela, sans devoir solliciter des gens. La quantité de données enregistrées étant très conséquente et les actions des utilisateurs ayant pu être capturées jusque dans leurs

moindres détails, il s’agit donc surtout de savoir quelles questions poser et comment des réponses peuvent y être apportées.

Trois niveaux d’analyse

Pour commencer, différents niveaux d’analyse peuvent être distingués. Il y a d’abord les fonctionnalités de base, c’est-à-dire tout ce que l’outil d’analyse, par exemple Piwik ou Google Analytics, permet de réaliser rapidement et facilement, sans que cela nécessite de compétences ou d’efforts particuliers. Il s’agit d’options “clés en main”. Ensuite, viennent les fonctionnalités avancées. Au sein de cette même interface, il existe des options qui permettent d’aller plus loin dans l’analyse et de répondre à des questions plus complexes. Cela implique toutefois plus de temps et d’efforts car il faut personnaliser l’outil en fonction du site web à analyser et de ce que l’on souhaite savoir. Ces interfaces d’analyse ont toutefois leurs limites. Comme expliqué dans la section précédente, il est parfois nécessaire d’utiliser les données brutes pour obtenir certaines informations. S’il est peu fréquent que le personnel des institutions ait du temps pour analyser en profondeur le trafic en ligne, le projet MADDLAIN a permis d’approfondir la recherche en explorant ce type de solutions plus artisanales, qui requièrent davantage de temps et sont techniquement plus exigeantes.

Les fonctionnalités de base permettent de répondre rapidement à des questions concernant les profils des visiteurs. Il est par exemple possible de voir sur une carte géographique d’où sont réalisées les connexions au site ou encore l’appareil utilisé pour se connecter (ordinateur, tablette ou smartphone).

En outre, dans le contexte du marketing web, il est fréquent de recourir à une série d’indicateurs-clés de performance, afin d’évaluer ce qui fonctionne bien et ce qui ne fonctionne pas sur un site web. Piwik propose par exemple des mesures d’engagement telles que la durée de la visite ou encore le nombre de pages vues.

Méthode des “goals”

La méthode des « goals », inspirée du secteur commercial, offre un niveau supplémentaire d’analyse. Il s’agit en effet de fonctionnalités servant à répondre à des questions plus complexes comme par exemple le pourcentage de visiteurs qui consultent des infos pratiques au cours de leurs visite, toutes pages confondues. Cette approche s’inscrit dans la continuité

du travail de Fagan¹². Les chercheurs du projet MADDLAN l'ont reprise et développée en construisant une méthodologie autour de six étapes indispensables :

1) Définition des goals

La première étape consiste à définir des goals, en répondant à ce type de questions : à quoi sert tel site web et qu'est-ce que le visiteur peut y trouver ? Qu'est-ce qui caractériserait une visite "réussie" ? Quelles en seraient les principales étapes ? Sur base d'une réflexion menée avec le personnel des institutions et de l'exploration attentive des différents sites web, une dizaine de goals ont été identifiés, avec quelques variantes en fonction des sites étudiés. Le nombre de goals s'avère être en effet un paramètre flexible qui doit être défini en fonction des besoins de l'institution et du temps disponible pour analyser les résultats obtenus. À titre illustratif, voici trois exemples de goals ayant été utilisés dans le cadre du projet¹³ : le goal "informations pratiques", qui se réfère à la consultation d'informations pratiques telles que tarifs, adresse ou heures d'ouverture ; le goal "aide et tutoriels", qui concerne toute l'aide en ligne : les FAQ, les tutoriels et manuels ; le goal "contenu numérisé", qui englobe la visualisation en ligne de documents numérisés.

2) Identification des données pertinentes

La liste de goals définie au cours de l'étape précédente permet d'explorer le site en profondeur et d'associer des pages web ou des actions (clics sur un lien ou download d'un fichier PDF par exemple) à chacun des goals. Ainsi, l'URL de cette page web sur les salles de lecture des Archives (<http://arch.arch.be/index.php?l=fr&m=nos-salles-de-lecture>) correspond par exemple au goal "informations pratiques" évoqué plus haut. Cela ne s'arrête toutefois pas là : il faut prendre en considération l'existence de trois autres différentes versions linguistiques pour cette page et répertorier quatre URLs différentes plutôt qu'une seule. Cette deuxième étape implique donc un travail rigoureux et méthodique, visant à prendre en considération l'ensemble des pages d'un site et toutes les traductions disponibles.

¹² Jody Condit Fagan, "The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment", in *Journal of Academic Librarianship*, 40(1), 2014, pp. 25-34.

¹³ Pour une description complète de cette méthode et une exhaustive des goals utilisés, voir : Jill Hungenaert et Anne Chardonens, "Combining Quantitative and Qualitative Methodologies to Analyse User Behaviour and Access Paths to Digital Content: Case Studies from Three Belgian Federal Scientific Institutions", in *Archives et Bibliothèques de Belgique*, 2017.

3) Création de filtres

Étant donné qu’il ne s’agit pas de consulter la popularité d’une page web isolée mais de créer une mesure pour chaque goal (impliquant potentiellement plusieurs pages web), il est nécessaire de rassembler les URLs collectées au cours de l’étape précédente au sein d’une seule expression. L’utilisation d’expressions régulières permet ainsi de créer un filtre prenant en compte l’ensemble des URLs¹⁴. Un exemple restreint et simplifié (fig. 4) offre un aperçu de ce processus. La présence du symbole “ | ” permet d’intégrer plusieurs possibilités et d’indiquer à la machine ce qu’elle doit prendre en compte. Cette étape assez technique nécessite de connaître les règles de syntaxe d’expressions régulières. La complexité du processus dépend surtout du nombre d’URLs à “compresser” en une seule expression.

4) Mise en œuvre des goals

Cette étape vise à paramétrer le goal au sein de l’interface Piwik. Il s’agit principalement de le personnaliser, de lui donner un nom et d’encoder le “filtre” permettant de voir qui atteint le goal.

5) Exécution d’un contrôle qualité

Il est important de tester ensuite si les goals implémentés fonctionnent correctement. Des simulations de parcours incluant tous les goals permettent de suivre en direct, au sein de l’interface de Piwik, si chaque goal est bien activé à la visite d’une URL donnée.

6) Réalisation de l’analyse et de la segmentation

À ce stade, tout a été paramétré, vérifié, et les données sont mesurées à chaque nouvelle visite. Il est désormais possible de savoir quel est le pourcentage de visites au cours desquelles un certain goal est atteint. Obtenir cette information est déjà une fin en soi. Il est cependant possible d’aller plus loin en affinant les résultats obtenus.

En effet, la foule indistincte des visiteurs peut être divisée en segments de visiteurs partageant

¹⁴ Pour en savoir plus sur les expressions régulières, voir par exemple : Jan Goyvaerts et Steven Levithan, *Regular Expressions Cookbook*, Sebastopol, 2009, pp. 510.

les mêmes caractéristiques. Cela permet par exemple de différencier les visiteurs selon leur provenance : un premier segment regroupant toutes les personnes qui se sont connectées directement sur le site, un deuxième celles qui sont arrivées via un moteur de recherche, un troisième celles qui ont cliqué sur le lien de la newsletter de l'institution et un quatrième celles qui ont suivi un lien présent sur un autre site web. Ces segments peuvent être utilisés pour vérifier des hypothèses et voir si certains comportements changent en fonction d'une caractéristique donnée, par exemple : est-ce que les personnes qui sont arrivées sur le site après avoir fait une recherche Google sont plus ou moins enclines à cliquer sur le lien vers le catalogue en ligne des archives ou de la bibliothèque ?

Solutions sur mesure

Comme expliqué dans la section concernant la collecte des données, certaines informations ne peuvent pas être obtenues en faisant uniquement appel à l'interface Piwik. Cela arrive notamment lorsque les URLs sont "ambiguës", c'est-à-dire qu'elles ne reflètent pas en détail l'activité du visiteur sur le site à moins qu'elles ne soient "augmentées" à l'aide de variables personnalisées. C'est le cas pour les URLs de Pallas par exemple, le catalogue numérique du CegeSoma. Or, le projet MADDLAIN vise notamment à évaluer l'utilisation du catalogue par le grand public à l'aide de données quantitatives. Il faut donc directement extraire l'information des données brutes pour savoir quelles sont les requêtes entrées par les utilisateurs dans le catalogue Pallas ou la façon dont les filtres de recherche sont utilisés. Les données utiles à l'analyse peuvent être extraites des données brutes stockées par le serveur à l'aide de commandes SQL. Une fois que les informations recherchées ont été localisées, elles peuvent être isolées dans un fichier structuré en vue d'être analysée de façon semi-automatisée, par exemple à l'aide de scripts Python, de la librairie Pandas ou du logiciel OpenRefine¹⁵. Ce processus est facilement reproductible, il requiert néanmoins un certain savoir-faire technique et doit être adapté aux spécificités de chaque site Web. Des détails supplémentaires sont fournis dans l'introduction de la partie II.

¹⁵ Une description détaillée de cette méthode a été présentée lors de la conférence ISI 2017 : Anne Chardonnes et Simon Hengchen, "Text Mining for User Query Analysis: A 5-Step Method for Cultural Heritage Institutions", in *Proceedings of the 15th International Symposium on Information Science (ISI 2017) Berlin, Germany*, 13-15 mars 2017.



Vigilance

Après ce bref survol des méthodes utilisées pour analyser les traces d'usage laissés par les utilisateurs, il est nécessaire de souligner que ce type d'approches reste assez expérimentale et qu'un certain nombre de limites subsiste. Pour n'en citer qu'une, il est clair que les données restent approximatives et que les méthodes actuelles ne permettent pas d'avoir de certitudes concernant la personne qui est réellement en train de visiter un site. Prenons par exemple le cas d'un visiteur qui se connecte au site de la KBR le matin depuis son smartphone afin de vérifier les heures d'ouverture. Celui-ci se connecte ensuite depuis son ordinateur portable au travail. Puis lors d'un travail de groupe, il navigue sur le site en compagnie de ses collègues, depuis un autre ordinateur. Enfin, en rentrant à la maison, ayant besoin d'une information urgente, il se connecte une dernière fois, depuis son ordinateur personnel. Pour Piwik, il s'agira à chaque fois de visites et de visiteurs différents. Par ailleurs, tous les collègues réunis derrière le même ordinateur apparaîtront comme un seul et "unique" visiteur. Ce type de limites doit être soigneusement pris en considération lors de la consultation des résultats. Cela ne signifie pas pour autant que toute analyse est vaine ou non valide mais nous rappelle simplement que les données obtenues par ce biais sont avant tout utiles pour identifier de grandes tendances, discerner des motifs récurrents ou encore nous rendre attentifs à certains phénomènes touchant un segment de visiteurs en particuliers.

Partie I : sites web institutionnels



Introduction

Cette première partie vise à permettre une meilleure connaissance des utilisateurs ainsi qu'une évaluation de la performance des sites web classique des institutions. Ces sites constituent généralement la première « porte d'entrée » vers les institutions et leurs collections. Les visiteurs peuvent notamment y trouver des informations pratiques, quelques mots sur les missions et l'histoire des institutions, des actualités, ainsi qu'un accès vers le catalogue et d'autres ressources en ligne (bibliothèque numérique, etc.).

L'une des particularités du projet MADDLAIN fut de rassembler au sein d'un même projet trois établissements scientifiques afin de pouvoir réaliser une analyse transversale sur les comportements et besoins de leurs utilisateurs. Ce fut donc l'occasion de confronter des données provenant de différents sites web. Bien que ceux-ci possèdent leur architecture propre, un nombre de visiteurs distinct ainsi que de nombreuses spécificités uniques, il fut néanmoins instructif de pouvoir comparer des pourcentages tels que le « bounce rate » (taux de visiteurs quittant le site directement, après n'avoir vu qu'une seule page). C'est ce type de données qui sera comparé au sein de cette section.

Étant donné que certaines pourcentages se prêtent particulièrement à une vue comparative et que d'autres nombres sont issus d'une analyse « sur mesure » pour l'un ou l'autre site, cette première partie se décompose en deux volets :

- le premier volet (1.1) propose une vue comparative pour les trois sites « portes d'entrée » des institutions (www.arch.be, www.kbr.be et cegesoma.be)
- le second volet (1.2) propose des analyses spécifiques pour chacun de ces trois sites.

Les sites web des catalogues en ligne, les sites d'accès aux collections numérisées et les sites de valorisation des collections, feront quant à eux l'objet de la seconde partie de ce rapport.

Précisons enfin que tous les résultats rassemblés dans cette section proviennent d'un échantillon d'analyse. Ce dernier est constitué de toutes les données (de fréquentation des sites) qui ont été collectées entre le **01.01.2016** et le **01.07.2016**.

1.1. Vue comparative : sites web institutionnels

Visites et visiteurs

Chaque fois qu'un visiteur vient sur un site pour la première fois, ou qu'il visite une page plus de 30 minutes¹⁶ après sa dernière page vue, une nouvelle visite est enregistrée. Un « visiteur » unique peut donc réaliser plusieurs visites par jour. À sa première visite, le visiteur est considéré comme un « New Visitor ». Dès sa seconde visite, il sera considéré comme un « Returning visitor ». Un même visiteur est reconnu grâce à l'usage de « cookies ». Malheureusement, Piwik ne nous donne le nombre de visiteurs uniques que pour une période d'un mois maximum. Ce type de données n'est pas accessible pour de plus longues périodes.

| | AGR- www ¹⁷ | KBR- www ¹⁸ | CGS- www ¹⁹ |
|---------------------------|---------------------------|---------------------------|---------------------------|
| Visites JANVIER | 39 351 | 23 837 | 4 455 |
| Visiteurs uniques JANVIER | 21 113 | 17 447 | 3 221 |
| Visites FÉVRIER | 27 856 | 24 142 | 4 607 |
| Visiteurs uniques FÉVRIER | 16 100 | 17 821 | 3 343 |
| Visites MARS | 39 786 | 26 445 | 4 313 |
| Visiteurs uniques MARS | 21 250 | 19 124 | 3 150 |
| Visites AVRIL | 37 380 | 22 312 | 3 915 |
| Visiteurs uniques AVRIL | 19 984 | 16 410 | 2 828 |
| Visites MAI | 33 758 | 22 079 | 4 614 |
| Visiteurs uniques MAI | 18 895 | 16 483 | 3 145 |
| Visites JUIN | 32 460 | 17 901 | 3 316 |
| Visiteurs uniques JUIN | 18 102 | 13 268 | 2 388 |

¹⁶ Cette durée, fixée par défaut à 30 minutes, peut être modifiée au sein de Piwik.

¹⁷ <http://www.arch.be>

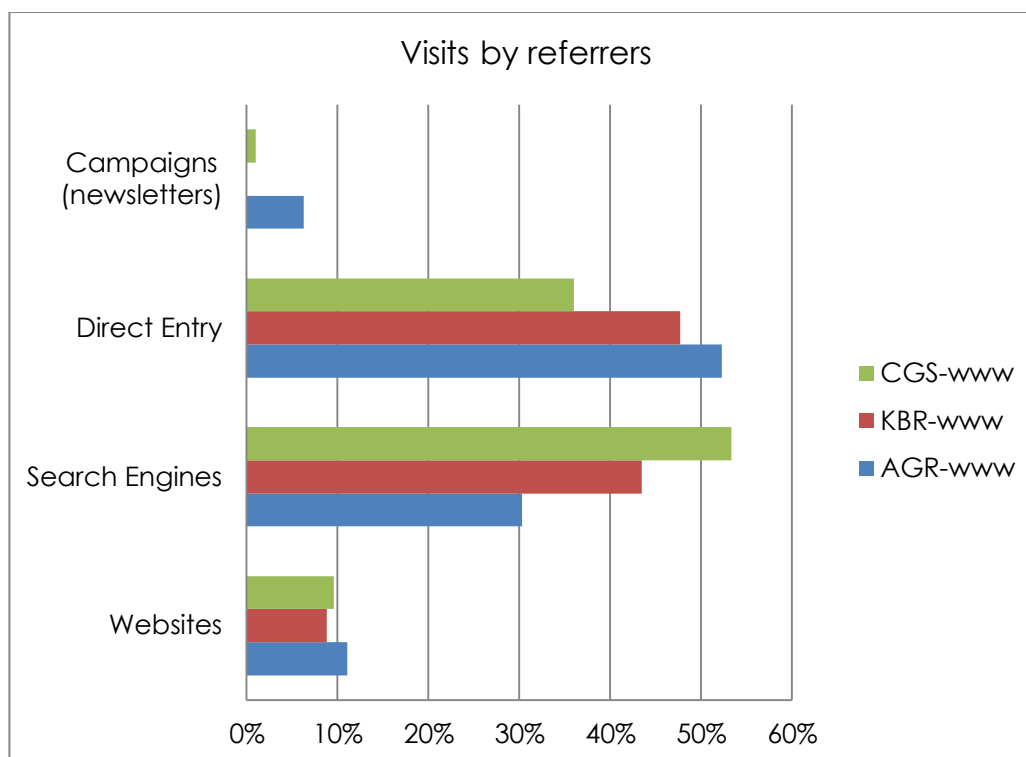
¹⁸ <http://www.kbr.be>

¹⁹ <http://www.cegesoma.be>

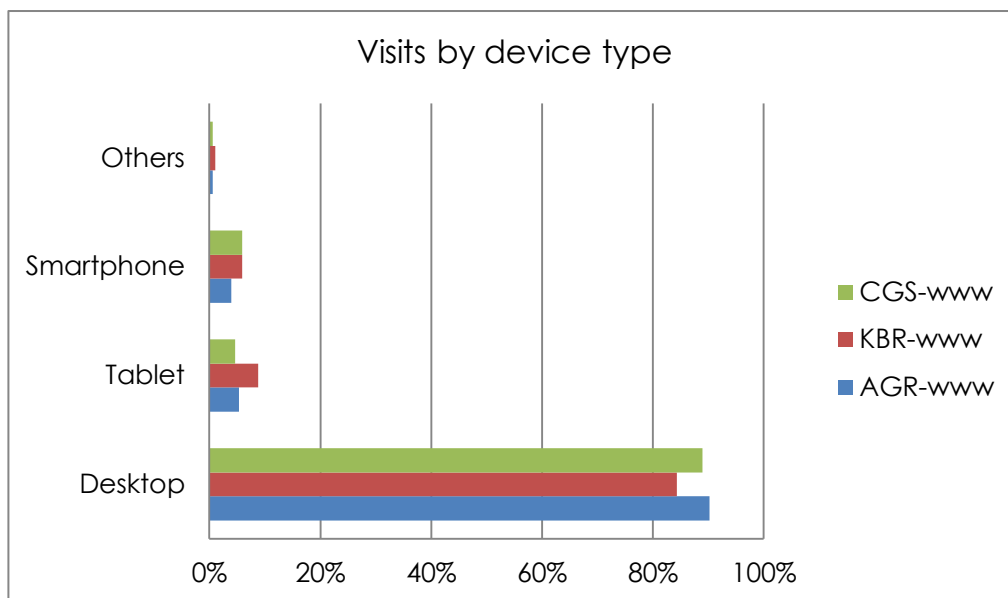
Répartition des visites

Les visites peuvent également être analysées de façon plus détaillée en les visualisant par segments. Par exemple, il est possible de connaître l'origine du trafic sur le site (on parle de « referrers ») et de répartir les visites selon cette variable-là. Il faut cependant noter que ce classement est réalisé automatiquement par Piwik. Ainsi, le segment « campaigns », qui peut par exemple prendre en compte les visites générées par un clic sur une newsletter, ne comporte aucune visite pour la KBR. Celle-ci envoie pourtant des newsletters renvoyant vers le site www.kbr.be. En fait, les liens n'ont pas été paramétrés pour que Piwik reconnaisse l'origine de ces clics. Il en va un peu de même pour le site du Cegesoma. Des visites sur www.cegesoma.be sont reprises dans la catégorie « campaigns », mais en réalité ces dernières proviennent de la newsletter envoyée par les Archives de l'État.

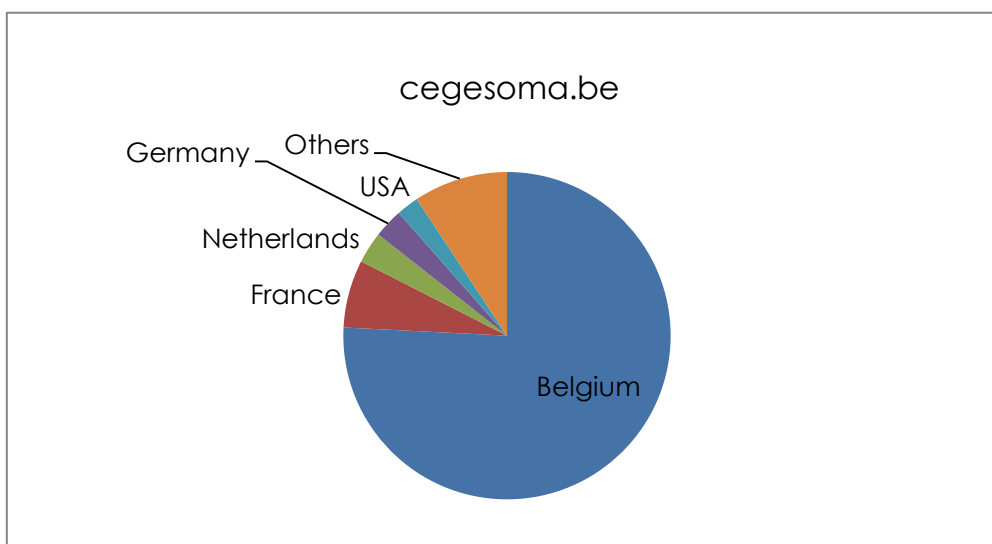
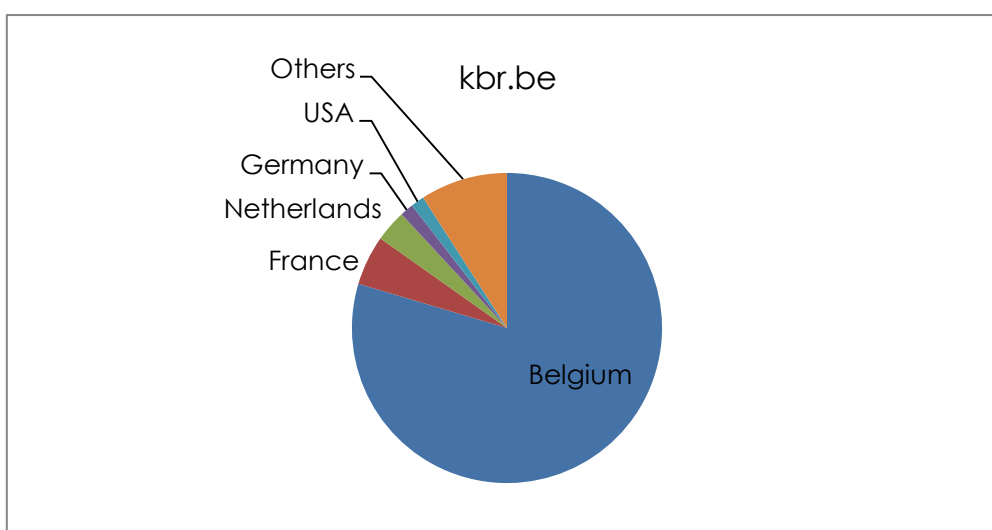
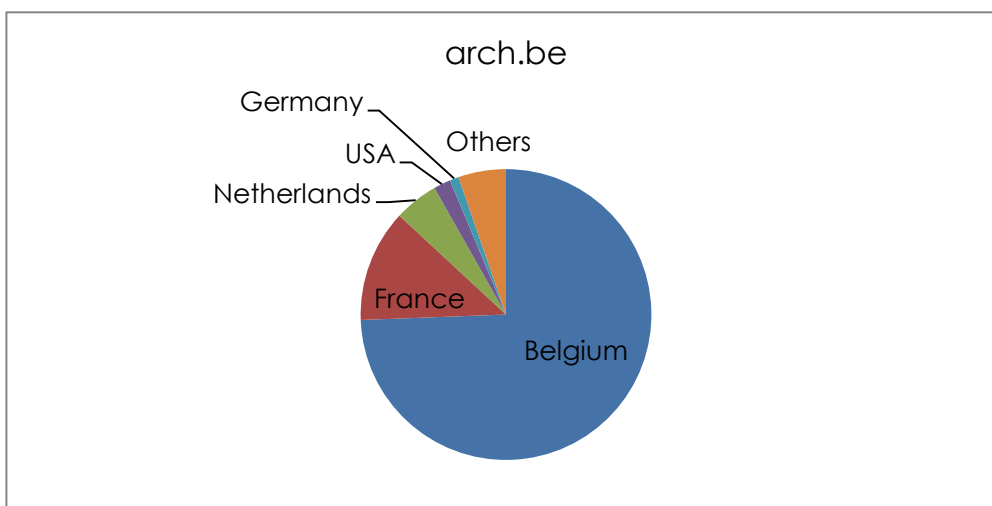
Répartition des visites par type d'entrée sur le site



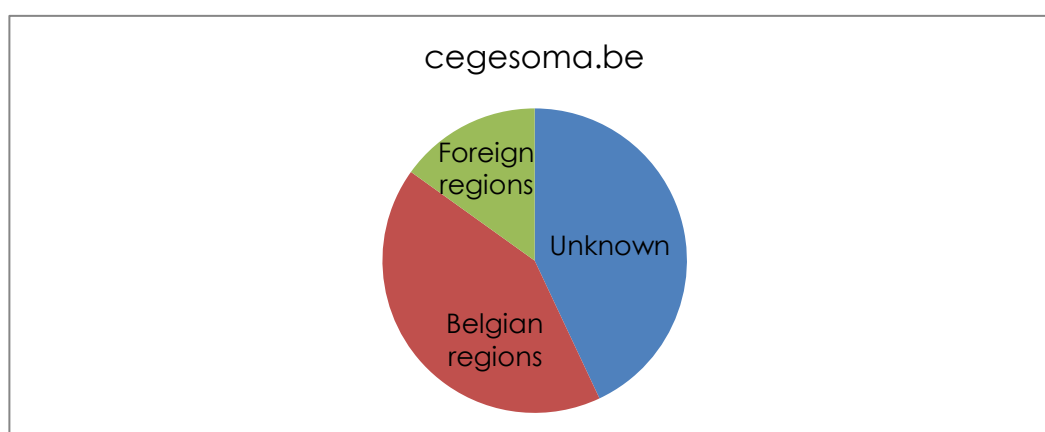
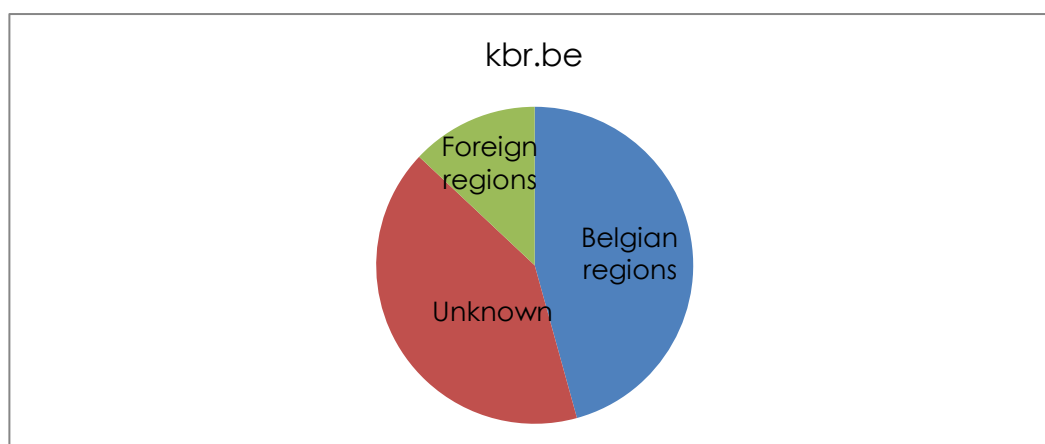
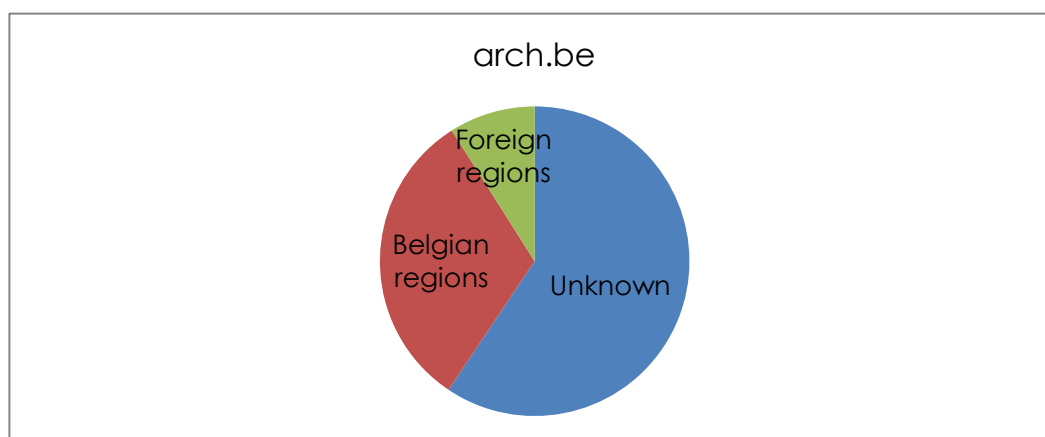
Répartition des visites par type de supports



Répartition des visites par pays

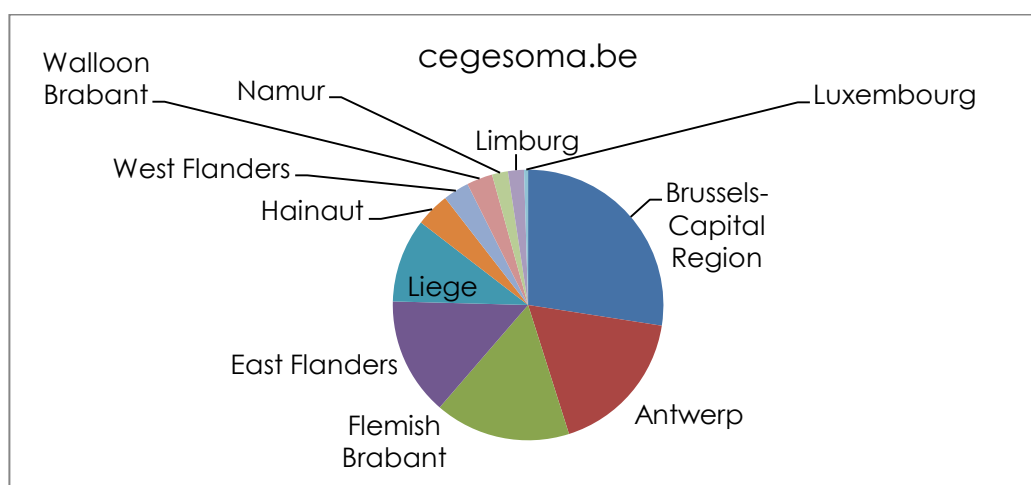
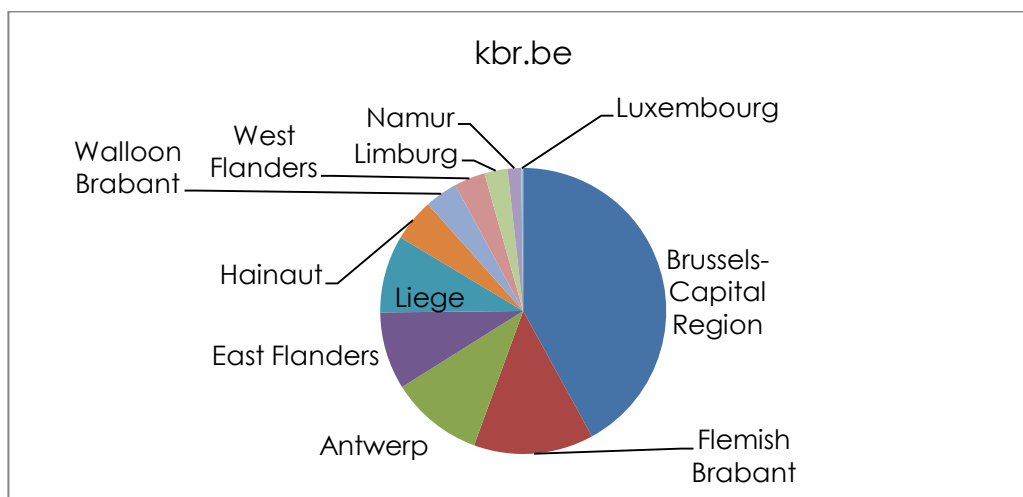
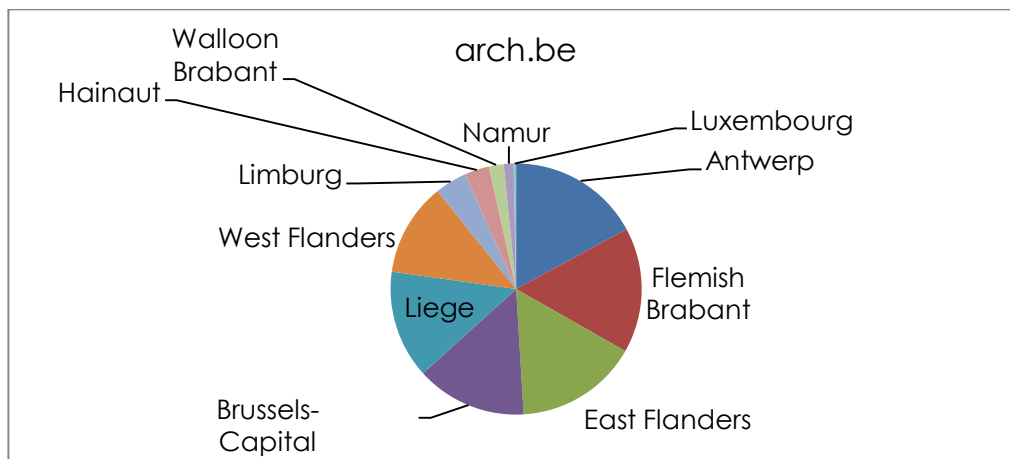


Répartition des visites par région²⁰



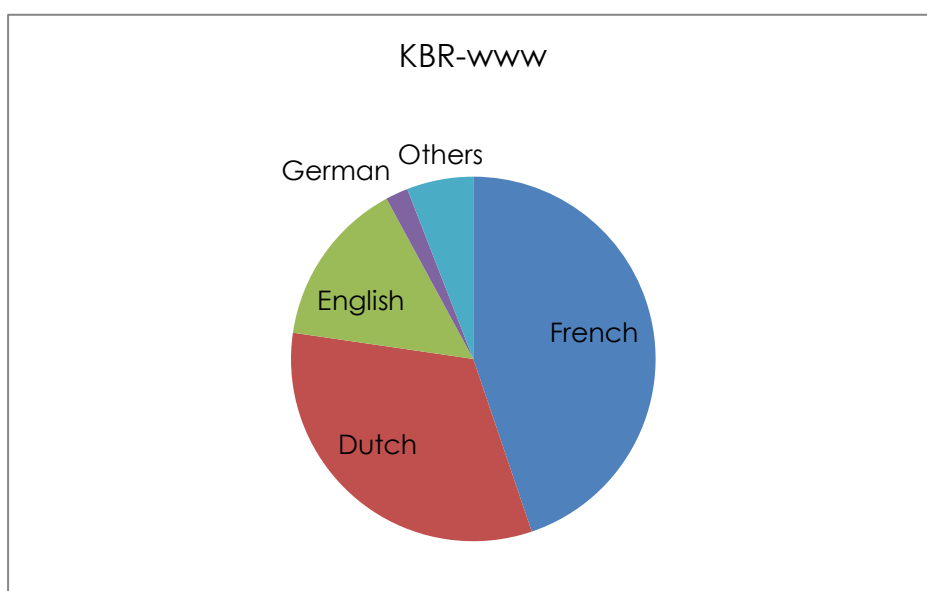
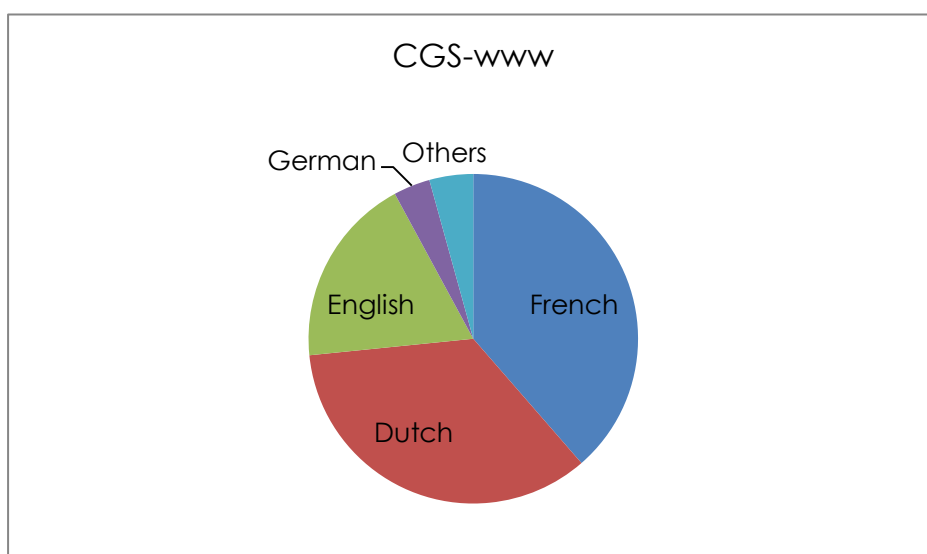
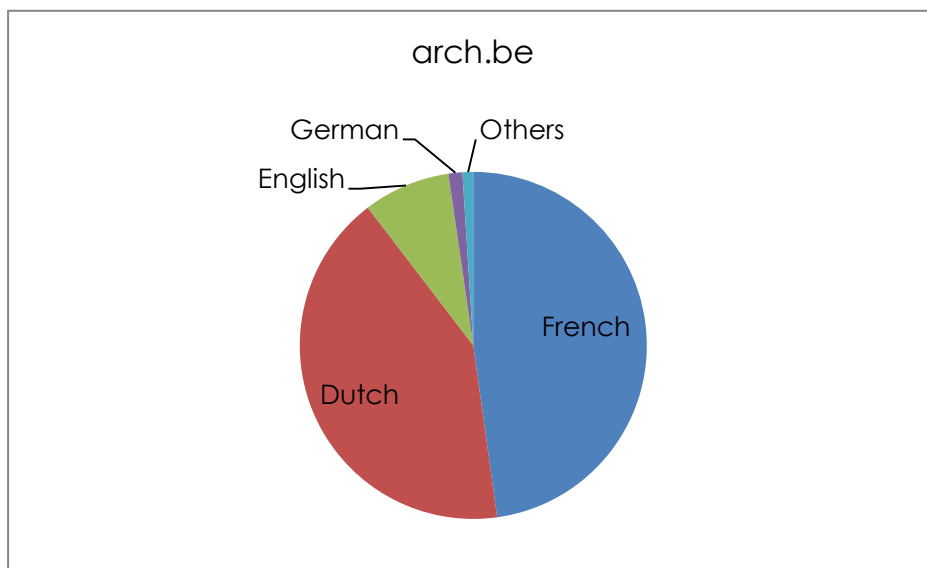
²⁰ Pour déterminer où une adresse IP est située géographiquement (géolocalisation), la base de données free GeoLiteCity a été utilisée, tel que recommandé par Piwik. Malheureusement, certains visiteurs ne peuvent pas être localisés (« unknown » label).

Répartition des visites issues de Belgique, par région²¹



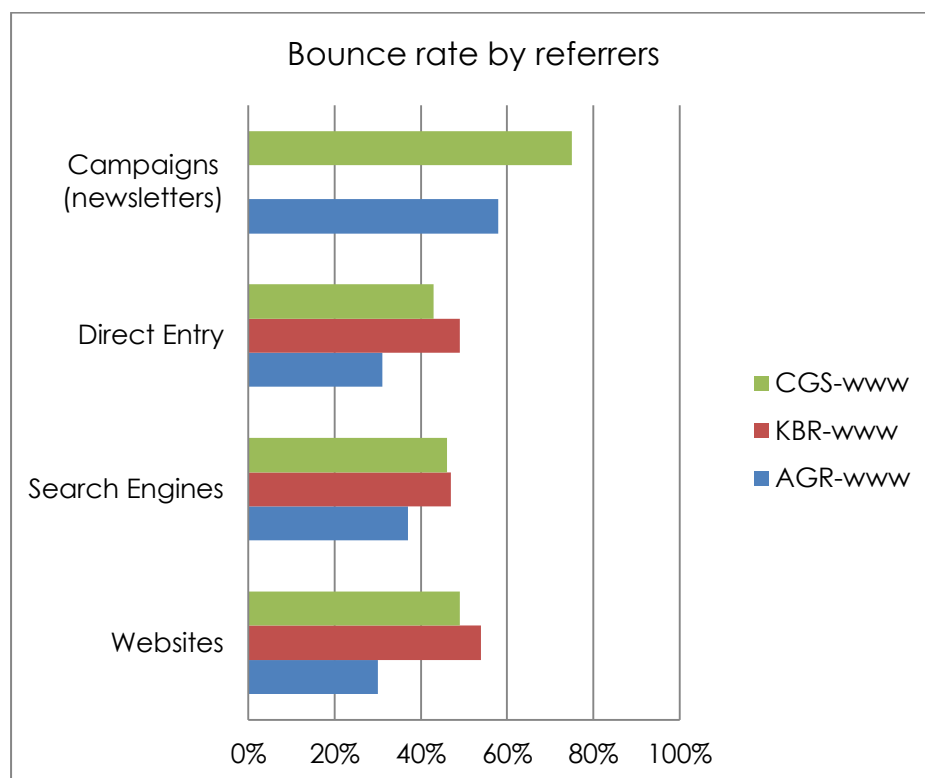
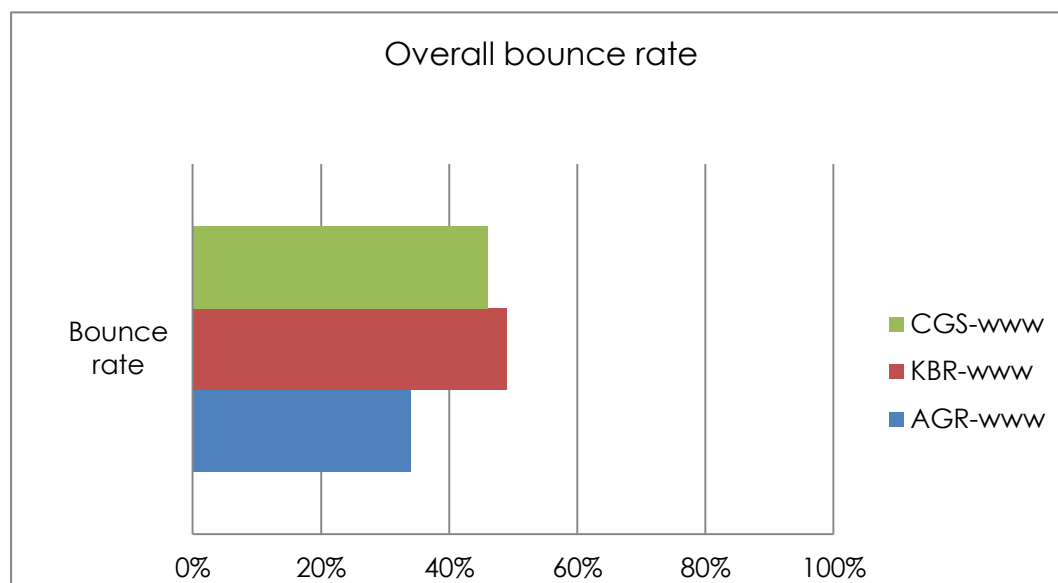
²¹ Tenir compte des « unknown » montrés dans les graphiques précédents.

Répartition des visites selon la langue du navigateur web



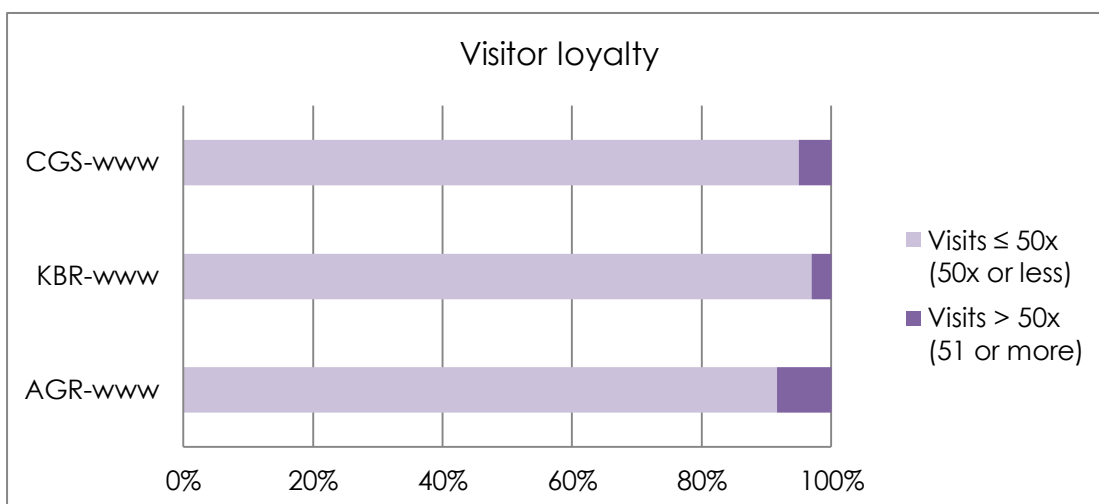
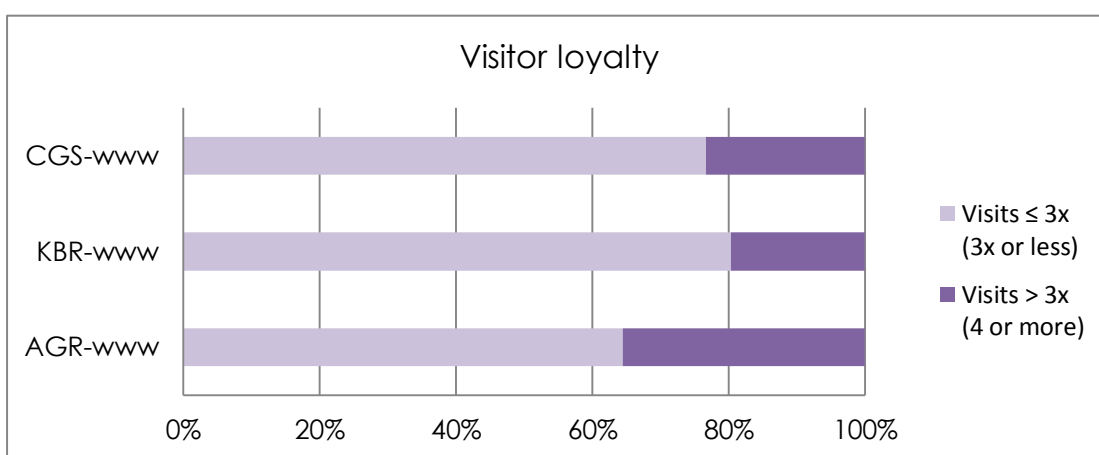
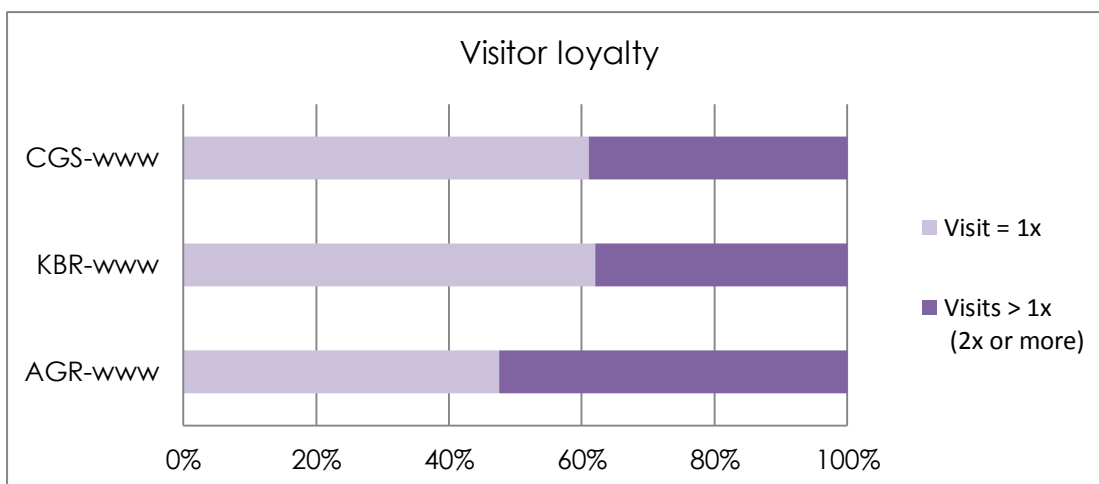
Bounce rate

Il s'agit du taux de visiteurs quittant le site directement, après n'avoir vu qu'une seule page : cela peut être la page d'accueil du site comme une autre page.



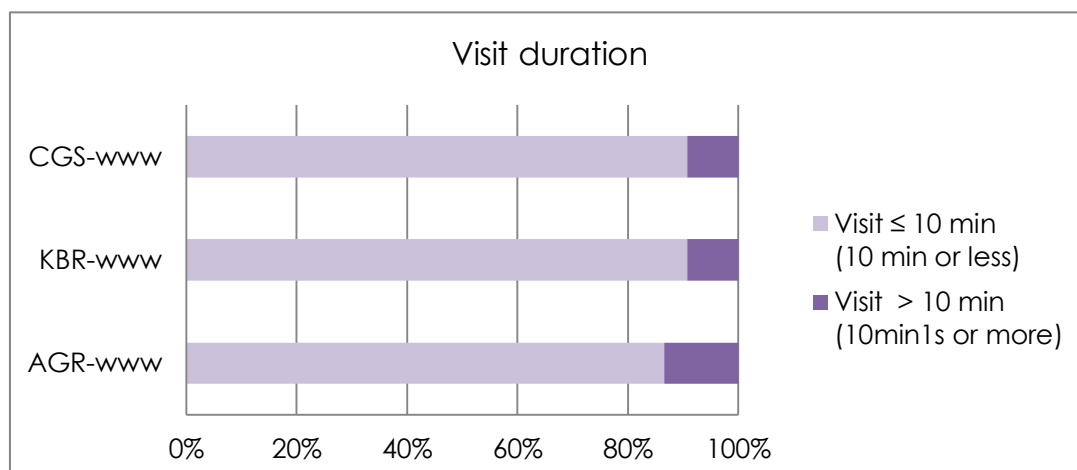
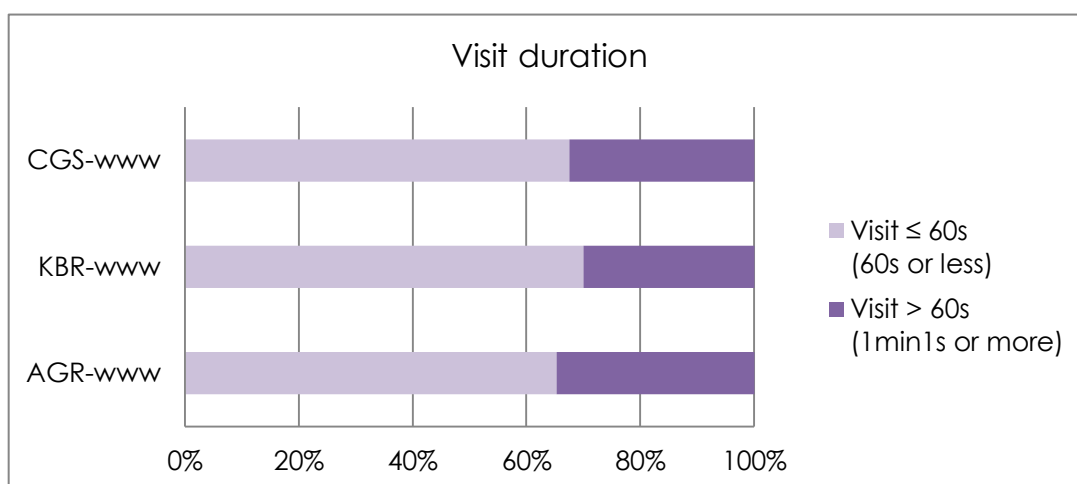
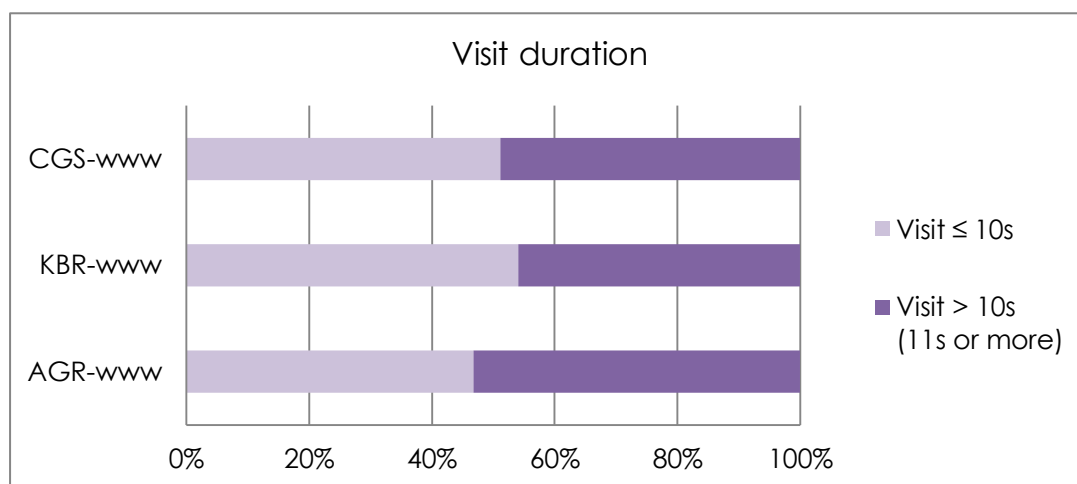
Visitor loyalty

Ce pourcentage permet de savoir si les visiteurs reviennent sur le site. On le calcule grâce au nombre de visites effectuées par une même personne durant la période d'analyse. On peut ainsi savoir, sur tous les visiteurs (100%), combien sont venus plus d'une fois.



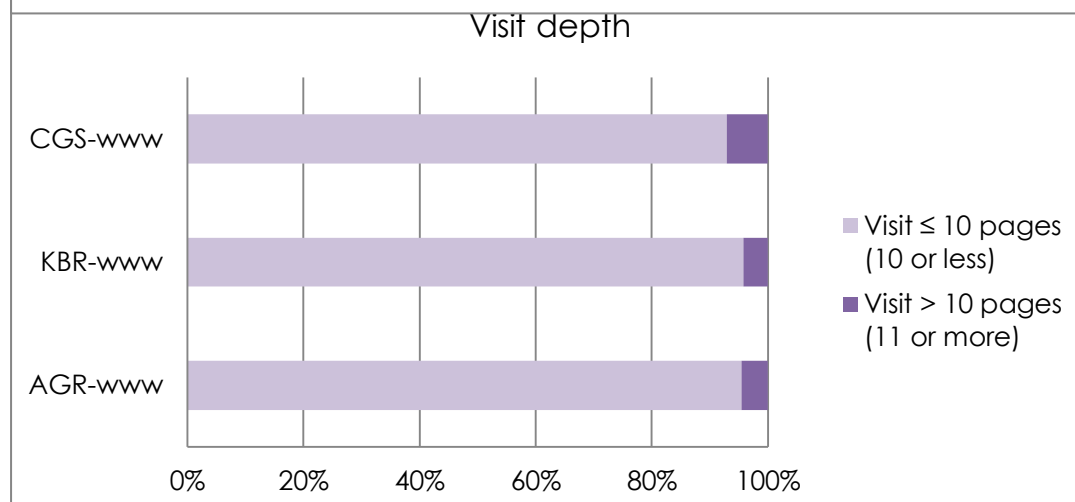
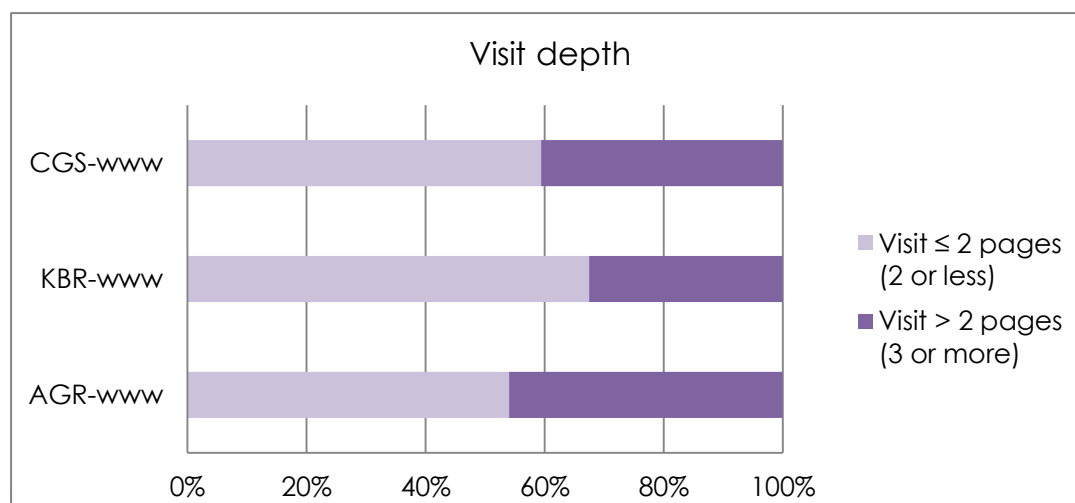
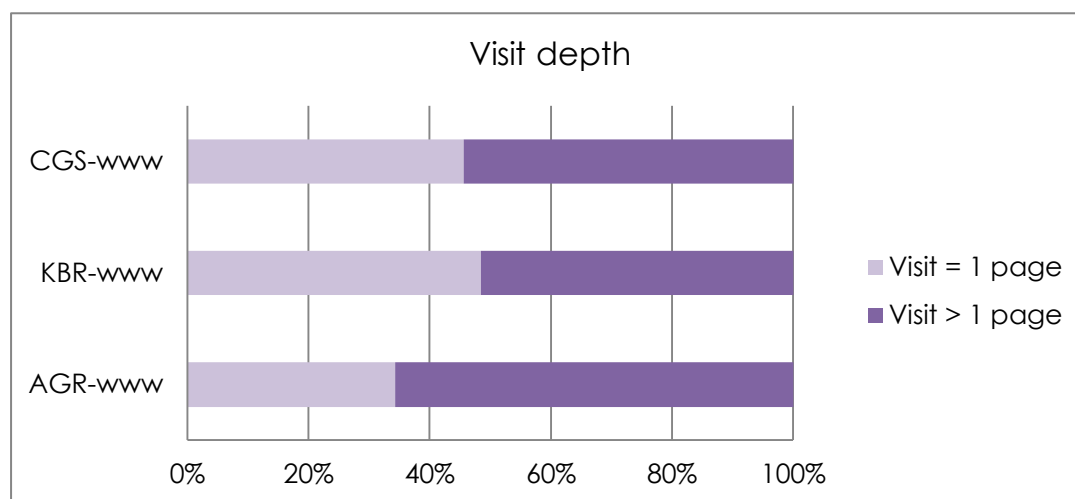
Visit duration

Il s'agit de prendre en compte la durée de chaque visite individuelle, afin d'obtenir des pourcentages globaux. Cela permet d'obtenir des infos plus précises qu'avec la plus classique durée moyenne par visite.



Visit depth

Il s'agit de connaître la profondeur des visites, c'est-à-dire le nombre de pages vues par visite. Cela permet par exemple de voir, ici sous forme de pourcentages, combien de visites comptent plus de trois pages vues.



1.2. Analyses spécifiques : sites web institutionnels

Cette partie du rapport consiste en une analyse approfondie du site web principal de chaque institution. En ce qui concerne le nombre de visites, le profil des visiteurs et leurs habitudes de visites, la section précédente présentait déjà un certain nombre d'informations. Nous les reprenons ici de façon synthétique et les complétons à l'aide de données supplémentaires.

Par ailleurs, comme expliqué plus haut (cf. Méthodologie), une part importante du travail de recherche a consisté à créer des goals sur mesure pour chaque site. Cela signifie que chaque analyse est unique et qu'elle dépend directement des caractéristiques du site concerné (contenu, volume, architecture de l'information, etc.).

Enfin, veuillez noter que si aucune information n'est donnée concernant la date, il s'agit toujours du même échantillon d'analyse couvrant une période de 6 mois (01.01.2016-01.07.2016).

1.2.1. Analyse spécifique : www.arch.be

The screenshot shows the homepage of the Rijksarchief in België. The header includes the logo and the text 'Rijksarchief in België' and 'ONS COLLECTIEF GEHEUGEN!'. The navigation menu is located below the header. The main content area is divided into several sections: 'IN DE KIJKER' with a photo of a building and text about publications; 'ONZE LEESZALEN' with a map of Belgium; 'SNEL ZOEKEN' with a search bar and radio buttons for different search options; and 'NIEUWS' with a news item about the 'Familiearchief Wittouck'.

COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 210 000 reprises. À noter qu'en raison d'un problème du serveur, seules 70 visites ont été comptabilisées entre le 6 et 14 février, ces nombres sont donc légèrement plus élevés en réalité.

Il est important de noter que parmi ces 210 000 visites, plus d'un tiers se sont terminées après que la visiteur n'a vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 34% pour le site web des Archives de l'État.

Par ailleurs, il nous est possible de savoir que plus de la moitié des visiteurs sont venus plus d'une fois sur le site (52,4%), plus d'un visiteur sur trois est venu quatre fois ou plus (35,5%) et 8,4% des visiteurs sont venus plus de 50 fois.

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que près de trois visites sur quatre (74,4%) sont effectuées depuis la Belgique, qui est suivie par la France (12,5%) et les Pays-Bas (4,9%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 59,4% dont on ne connaît pas l'emplacement exact) : Anvers (5,4%), le Brabant flamand (5,1%), la Flandre orientale (5,0%) et Bruxelles-Capitale (4,5%).

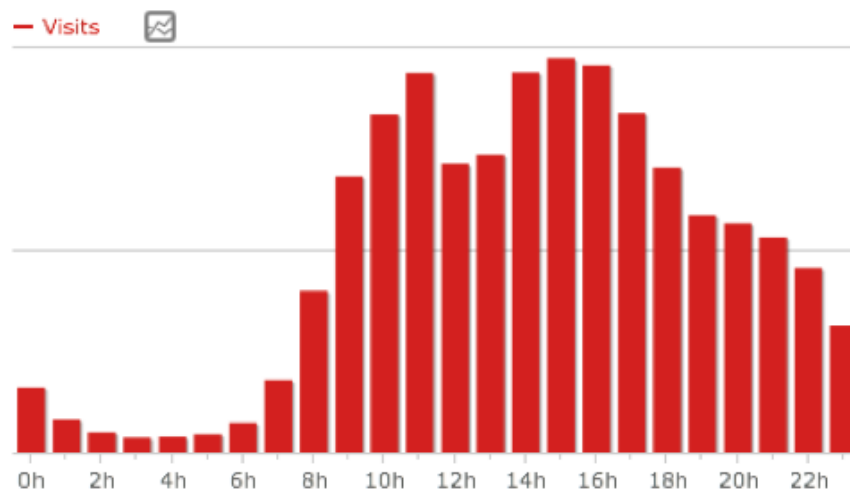
Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (47,8%), suivis de près par le néerlandais (41,8%), viennent ensuite l'anglais (8,1%) et l'allemand (1,3%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (90,2%). Parmi les autres modes de connexion, nous savons que 5,3% des visites sont réalisées à partir d'une tablette et 3,9% via une tablette.

QUELLES SONT LEURS HABITUDES DE VISITE ?

Parmi toutes les visites effectuées sur le site, plus de la moitié (53,2%) durent plus de 10 secondes. Un peu plus d'un tiers des visites (34,7%) durent plus d'une minute, 13,4% durent plus de 10 minutes et enfin 3,2% durent plus de 30 minutes.

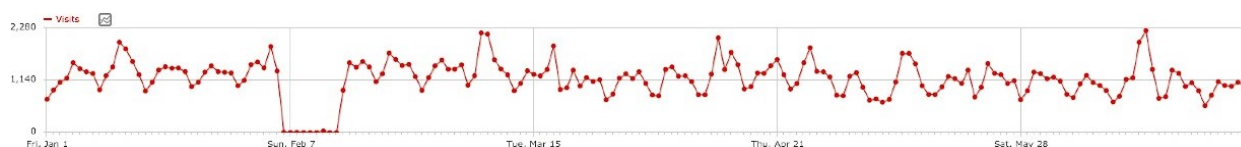
Nous pouvons également nous intéresser à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'un nombre de pages élevé peut être tant un signe de succès (site intéressant qui a su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information recherchée ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu). Pour le site www.arch.be, plus de 45 % des visites comptabilisent trois pages vues ou plus (46,0%), et 4,5% des visites comptent plus de 10 pages vues.



www.arch.be - Fréquentation par heure (selon l'heure du serveur)



Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 21h. De plus, nous pouvons également observer l'évolution des visites à travers le temps. On remarque la panne du serveur signalée précédemment et, de manière plus générale, nous observons une baisse de fréquentation régulière durant le week-end.



www.arch.be – Évolution des visites, jour par jour

Enfin, il est également possible de voir l'impact positif des newsletters de l'institution sur le taux de fréquentation du site. En effet, des pics d'affluence suivent chaque envoi. C'est le cas notamment le 16 juin (plus de 2200 visites), lendemain de l'envoi de la newsletter de juin (en néerlandais). Cette hypothèse est confirmée par le contenu des pages les plus vues ce jour-là, ainsi que par le type provenance des visiteurs.

D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour le [site Arch](#), nous savons que plus de la moitié des visites (52,3%) se font en accès direct, c'est-à-dire que la personne a soit enregistré le site dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Un tiers des visites (30,3%) sont le fruit d'une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduits au site web des Archives de l'État pourrait être très instructif. Malheureusement la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Nous avons ainsi plus de 60 000 mots-clés « not defined ». Les mots-clés disponibles ayant occasionné plus de 100 visites sont : « rijksarchief » (239 visites) ; « agr » (138 visites) et « arch.be » (114 visites)²².

²² Cela s'explique peut-être par le fait qu'aucune redirection automatique vers <http://www.arch.be> n'est prévue quand on écrit seulement « arch.be » dans son navigateur.

Plus de 10% des visites (11,1%) correspondent à un clic sur un lien vers le site Arch effectué sur d'autres sites web/réseaux sociaux. Nous retrouvons Facebook en tête de liste, cependant le détail est un peu confus et il n'est pas certain que tous les postes Facebook aient été pris en compte.

Enfin, la newsletter de l'institution est à l'origine de 6,3% des visites. Comme évoqué précédemment, on peut d'ailleurs observer des pics de fréquentation sur le site aux moments des envois.

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE / SORTIE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ces forces et ces faiblesses.

En ce qui concerne les pages d'entrées les plus fréquentes, nous trouvons en premier la page d'accueil en français (13,1%), la page d'accueil générale (12,1%), la page d'accueil en néerlandais (9,4%). Viennent ensuite les pages présentant les pages consacrées aux généalogistes, en français (5,1%) et en néerlandais (4,2%).

En ce qui concerne les pages ayant un taux de sortie élevé, nous trouvons exactement les mêmes pages (pages d'accueil dans les différentes langues + pages consacrées aux généalogistes). Ce constat est sujet à interprétation et il est difficile de déterminer précisément quelle est la logique à l'œuvre. Une hypothèse est que ce taux de sortie élevé pourrait être causé par deux facteurs distincts : le bounce rate (personnes quittant immédiatement le site) et les clics sur les liens renvoyant vers le catalogue en ligne (Search). Les statistiques disponibles semblent aller dans le sens de cette hypothèse, d'autres raisons sont toutefois possibles.

GOALS

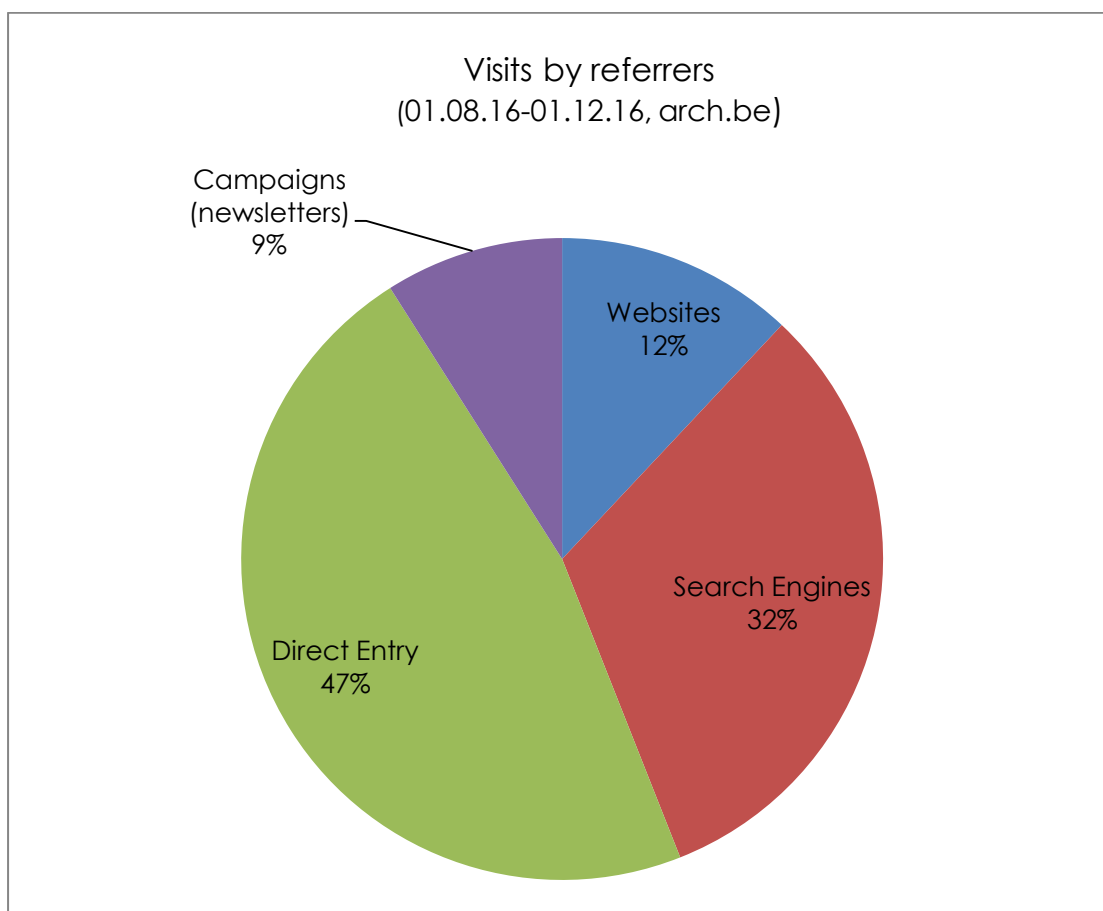
Pour le site arch.be, 10 goals ont été identifiés :

| ARCH.BE | | | |
|----------|-------------------------|--|----------------------------------|
| Name | Access to | Web pages that contain | Taux de conversion ²³ |
| PRA | Practical information | Contact info, address, opening hours, prices of reading room cards | 29,4% |
| CONT II | Meta-information | Historical background, institution's mission and structure | 18,0% |
| CONT III | News and agenda | Agenda, news, newsletter subscription | 20,7% |
| TUT | Help and tutorial | FAQ, tutorials, site and search engine manuals, help pages, search assistance | 4,9% |
| TUT-PDF | Help and tutorial - PDF | FAQ, tutorials, site and search engine manuals, help pages, search assistance (PDF) | 1,0% |
| VAL | Valorisation documents | Articles about collections | 3,8% |
| CAT | Main catalogue | Hyperlinks to main digital catalogue | 38,7% |
| REP | Ordering reproductions | Information related to ordering reproductions | 0,6% |
| PROFILES | Information by profile | <i>Additional goal, special « arch.be »</i> Personalised pages by profile of visitors (genealogist, researcher, etc.) | 28,3% |
| TRI | Record schedules | <i>Additional goal, special « arch.be »</i> Online resources : record schedules | 1,7% |

²³ Moyenne.

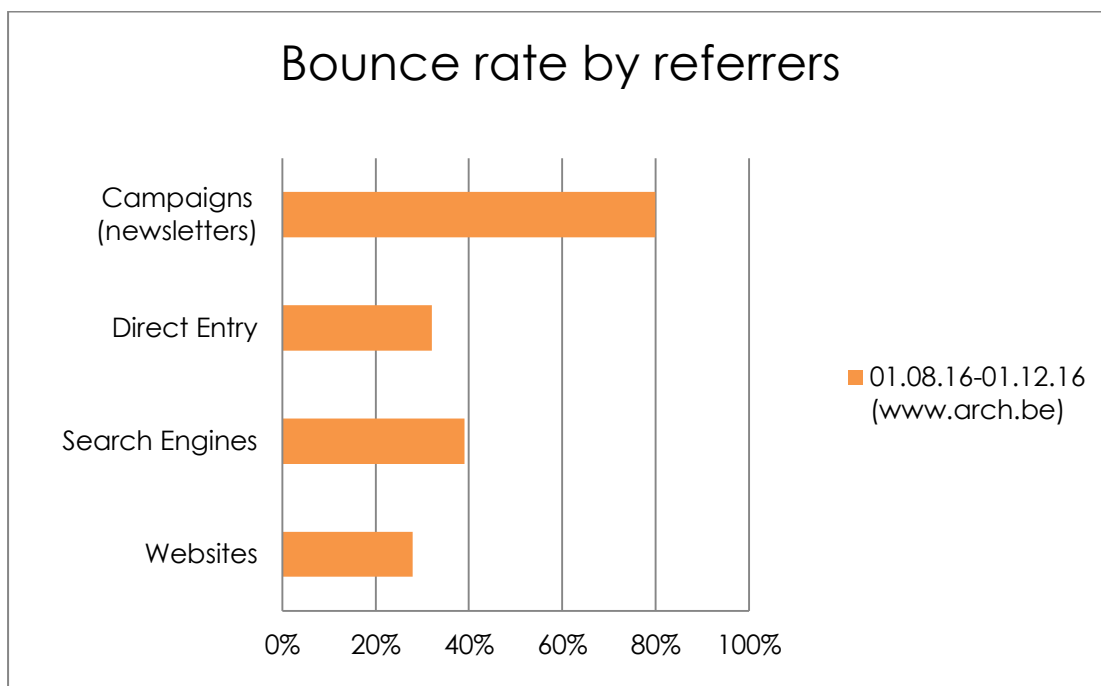
Les taux de conversion concernant le site arch.be proviennent tous d'une période d'analyse de 4 mois (01.08.2016-01.12.2016)²⁴.

Durant ce laps de temps, 146 249 visites ont été comptabilisées. Avant de présenter les taux de conversion par goal, il semble important de d'abord se pencher sur l'origine de ces visites. Segmentation des visites effectuées entre le 01.08.2016 et le 01.12.2016 :



²⁴ Les goals, qui ne sont pas rétroactifs, ont été mis en place au début du mois de juin 2016, au terme d'un long processus préparatoire. C'est pour cette raison que la période d'analyse n'est pas la même que pour la première partie du rapport.

Par ailleurs, si l'on veut se pencher sur les taux de conversion d'un site, il est utile de garder en tête le bounce rate de chaque segment (si l'on sait par exemple que plus de la moitié des utilisateurs quittent le site dès la page d'accueil, alors un taux de conversion de 20% semble directement plus encourageant).



PRA = informations pratiques = 29,4%

Cela signifie que 29,4% des visites (c'est-à-dire 43 031) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=nos-salles-de-lecture>

<http://www.arch.be/index.php?l=fr&m=contact>

<http://www.arch.be/index.php?l=fr&m=horaires>

<http://www.arch.be/index.php?l=fr&m=en-pratique&r=nos-salles-de-lecture>

<http://www.arch.be/index.php?l=fr&m=en-pratique&r=nos-salles-de-lecture&d=anderlecht>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=nos-salles-de-lecture>

<http://www.arch.be/index.php?l=fr&m=en-pratique&r=en-vente-aux-archives-de-l-etat>

<http://www.arch.be/index.php?l=fr&m=en-pratique&r=tarifs>

<http://www.arch.be/index.php?l=fr&m=enseignant&r=visites-guidees>

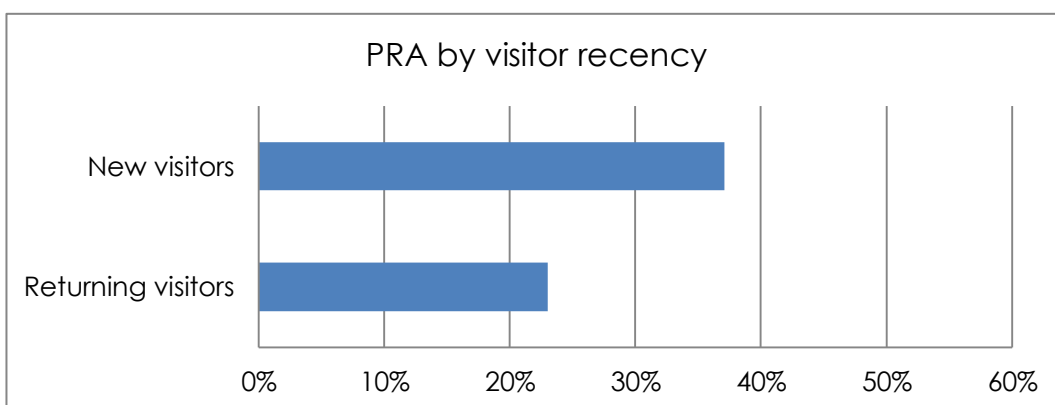
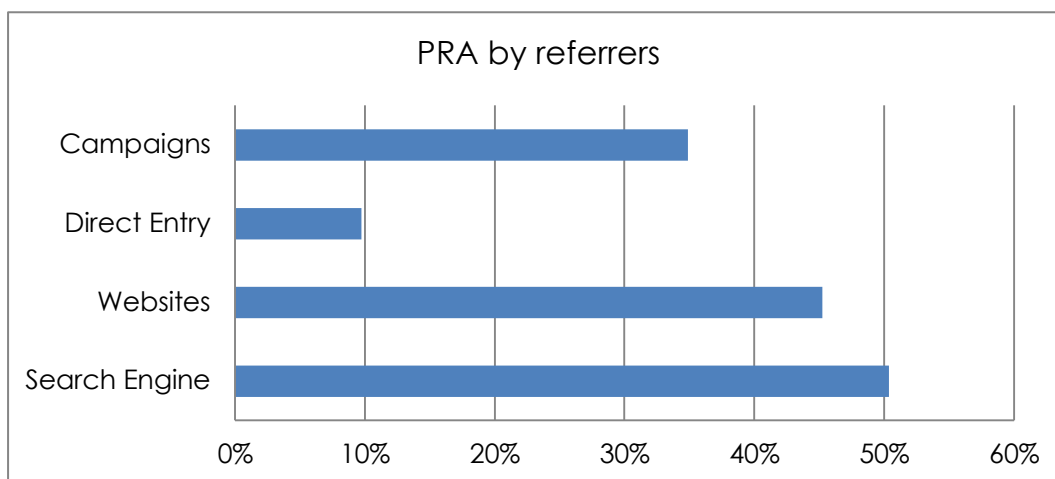
<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=contact-fonctionnaire>

<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=suivre-une-formation>

Etc.²⁵

²⁵ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires.

Taux de conversion par segment



CONT II = Meta-information = 18,0%

Cela signifie que 18,0% des visites (c'est-à-dire 26 349) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=l-institution&r=presentation>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=presentation#4>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=rapports-annuels>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=notre-role>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=que-conservons-nous>

<http://www.arch.be/index.php?l=fr&m=l-institution&r=historique>

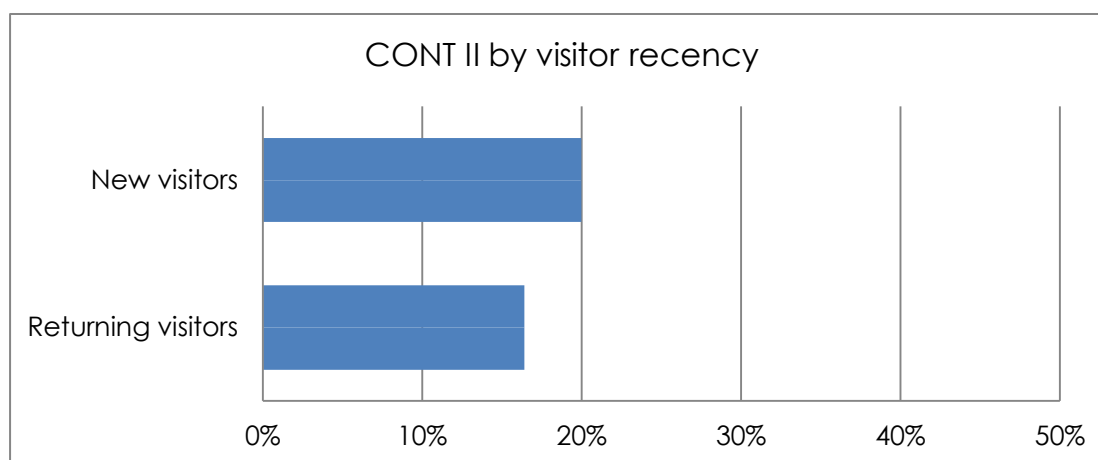
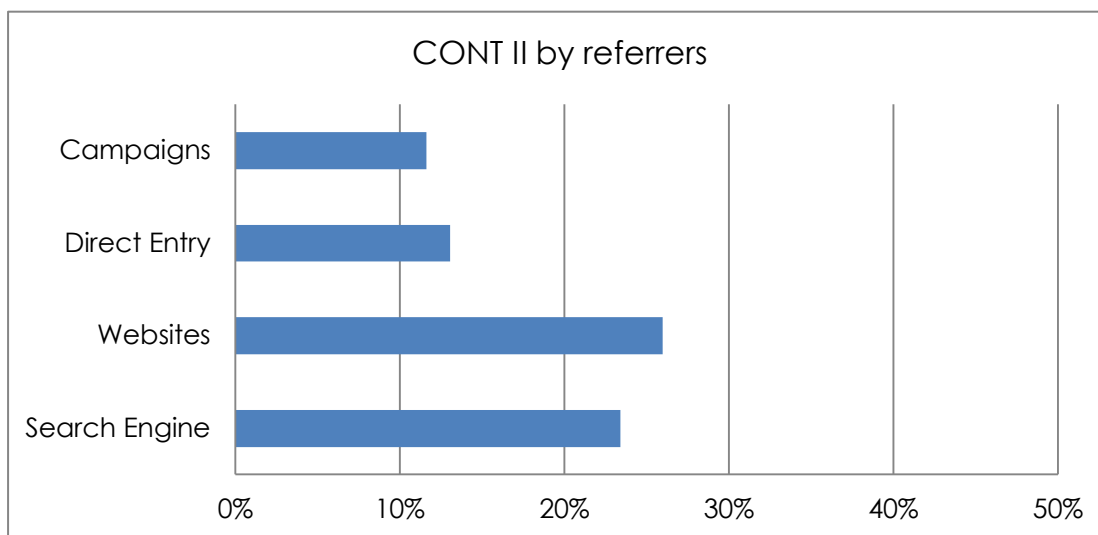
<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne>

<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=publications>

<http://www.arch.be/index.php?l=fr&m=nos-projets&r=projets-de-recherche>

<http://www.arch.be/index.php?l=en&m=our-projects&r=volunteer-projects>

Taux de conversion par segment²⁶



²⁶ Le recours à d'autres segments est bien sûr envisageable !

CONT III = Nouvelles et agenda = 20,7%

Cela signifie que 20,7% des visites (c'est-à-dire 30 219) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=bulletins-d-informations>

<http://www.arch.be/index.php?l=fr&m=actualites&r=toutes-les-actualites>

<http://www.arch.be/index.php?l=fr&m=actualites&r=agenda>

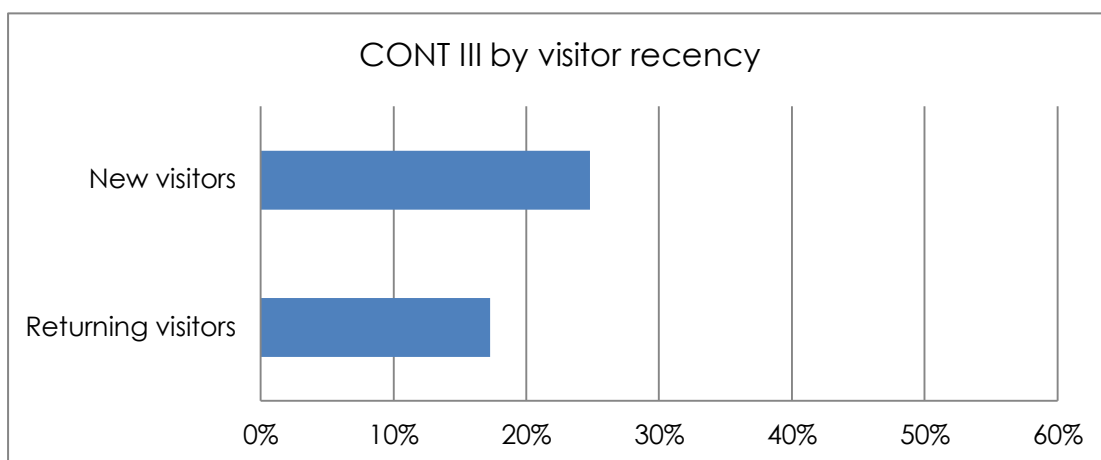
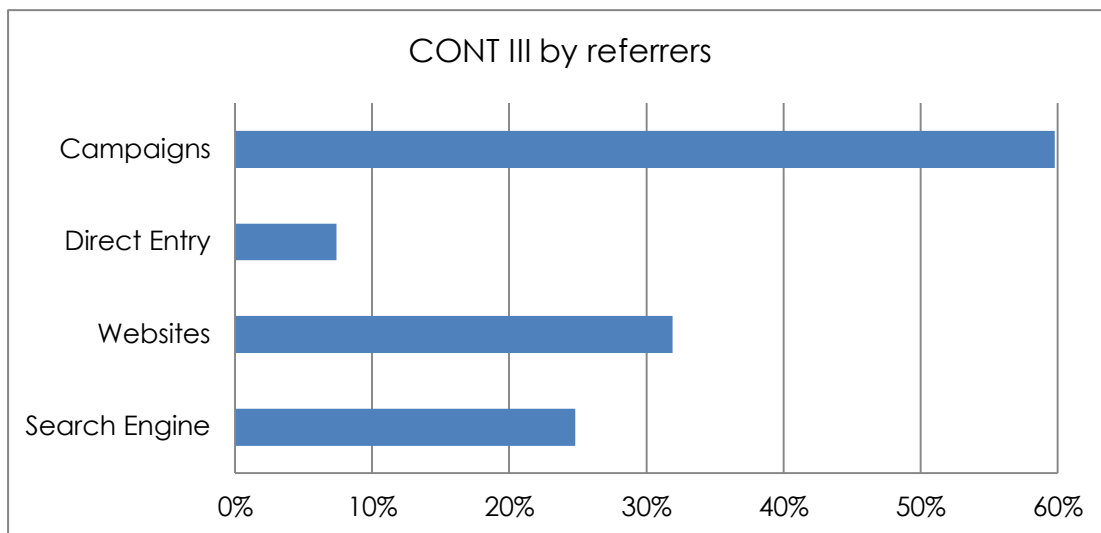
<http://www.arch.be/index.php?l=fr&m=actualites&r=expositions>

<http://www.arch.be/index.php?l=fr&m=actualites&r=journees-d-etude>

<http://www.arch.be/index.php?l=fr&m=actualites&r=conferences>

<http://www.arch.be/index.php?l=fr&m=actualites&r=reseaux-sociaux>

Taux de conversion par segment²⁷



²⁷ Le recours à d'autres segments est bien sûr envisageable !

TUT = Aide et tutoriels = 4,9%

Cela signifie que 4,9% des visites (7 179) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=aide>

<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=faq>

<http://www.arch.be/index.php?l=fr&m=chercheur&r=faq>

<http://www.arch.be/index.php?l=fr&m=genealogiste&r=faq>

<http://www.arch.be/index.php?l=fr&m=genealogiste&r=la-genealogie-par-ou-commencer>

<http://www.arch.be/index.php?l=fr&m=genealogiste&r=demogen>

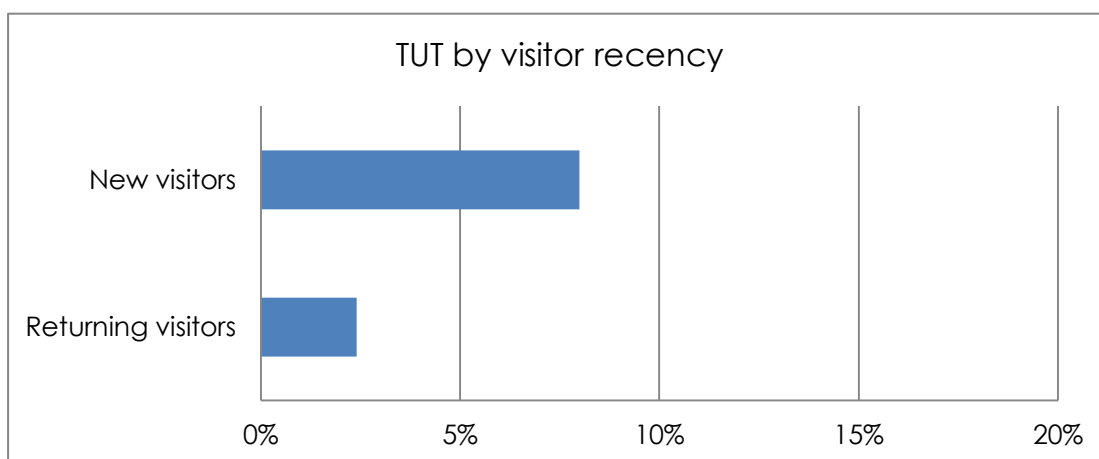
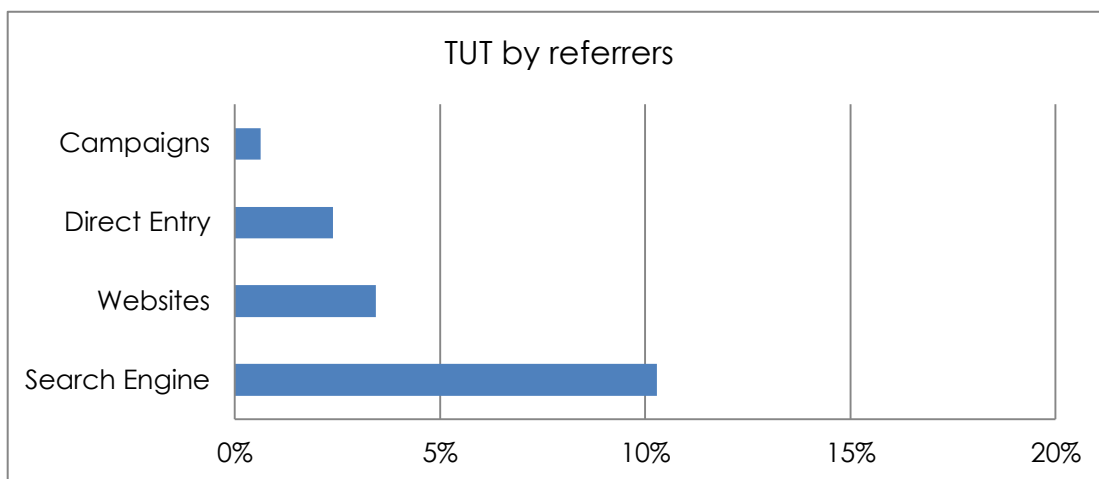
<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=faq-gestion-des-documents>

<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=faq-gestion-des-documents&p=elimination-d-archives>

<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=faq-gestion-des-documents&p=tri-d-archives>

<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=faq-gestion-des-documents&p=transfert-d-archives-aux-archives-de-l-etat>

Taux de conversion par segment

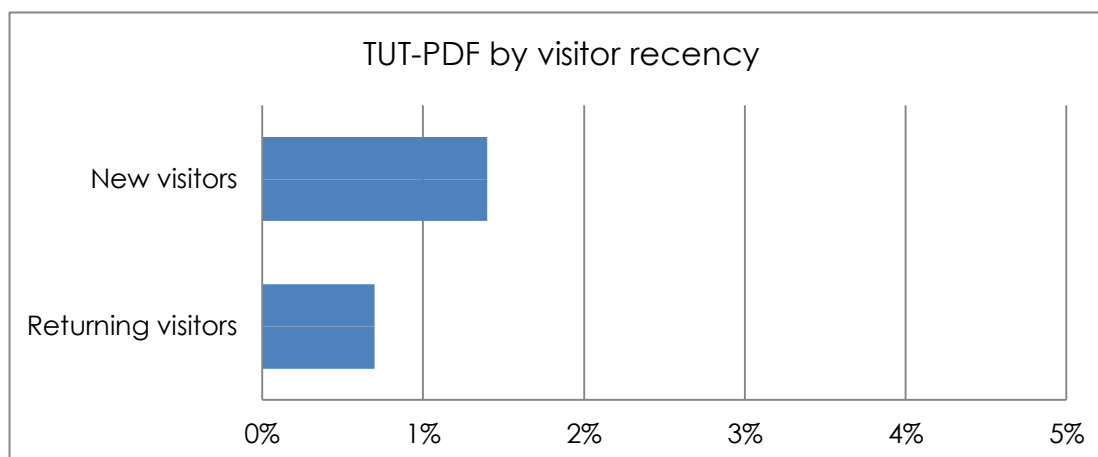
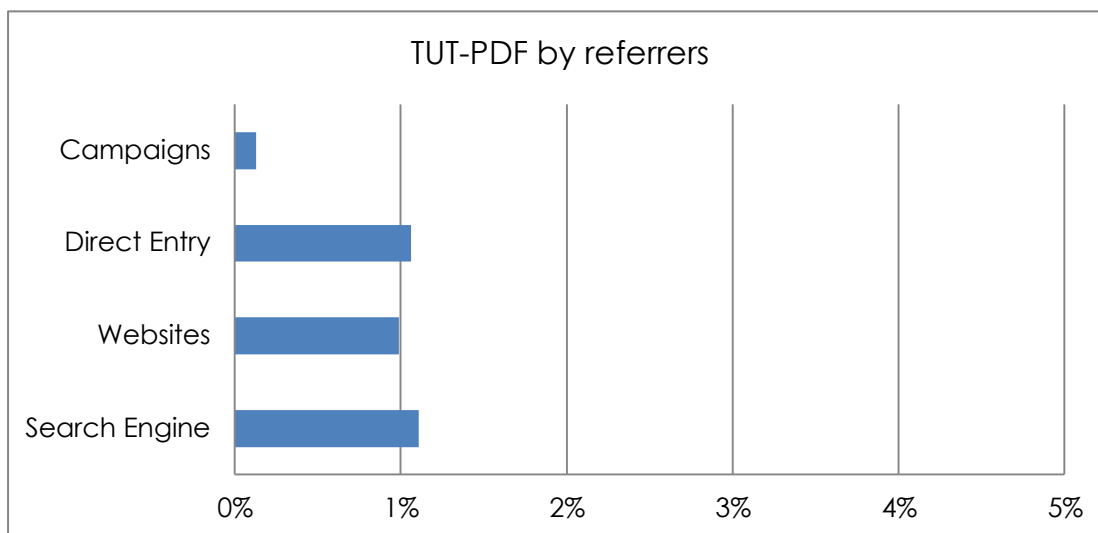


TUT-PDF = Aide et tutoriels pdf = 1,0%

Cela signifie que 1,0% des visites (1 440) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

http://www.arch.be/docs/brochures/rechercher_dans_les_archives_un_plan_par_etapes.pdf
http://www.arch.be/docs/brochures/notaire_notarissen/notaires_instructions_externes.pdf
http://www.arch.be/docs/brochures/notaire_notarissen/notaires_guide-depot-minutes-repertoires.pdf
http://www.arch.be/docs/brochures/transfert_archives.pdf
http://www.arch.be/docs/brochures/transfert_archives_numeriques.pdf
http://www.arch.be/docs/brochures/notaire_notarissen/notaires_instructions_externes.pdf
http://www.arch.be/docs/brochures/rechercher_dans_les_archives_un_plan_par_etapes.pdf
http://www.arch.be/docs/brochures/les_archives_cest_quoi.pdf
http://www.arch.be/docs/brochures/classer_documents.pdf
http://www.arch.be/docs/brochures/la_numerisation_en_marche.pdf
http://www.arch.be/docs/brochures/conserver_documents_numeriques.pdf
http://www.arch.be/docs/brochures/vous_demenagez.pdf
http://www.arch.be/docs/brochures/organisez_la_gestion_de_vos_archives.pdf
http://www.arch.be/docs/brochures/locaux_d_archives.pdf
http://www.arch.be/docs/brochures/depots_d_archives.pdf
http://www.arch.be/docs/brochures/archives_photographiques.pdf
http://www.arch.be/docs/brochures/transfert_archives.pdf
http://www.arch.be/docs/brochures/transfert_archives_numeriques.pdf
http://www.arch.be/docs/surv-toe/conseils_delais_utilite_administrative.doc
http://www.arch.be/docs/brochures/rediger_tableau_de_tri.pdf

Taux de conversion par segment



VAL = Valorisations (documents) = 3,8%

Cela signifie que 3,8% des visites (5 592) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=actualites&r=expositions&sr=expositions-virtuelles>

<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=expositions-virtuelles>

<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=moteurs-de-recherche&sr=bases-de-donnees-thematiques>

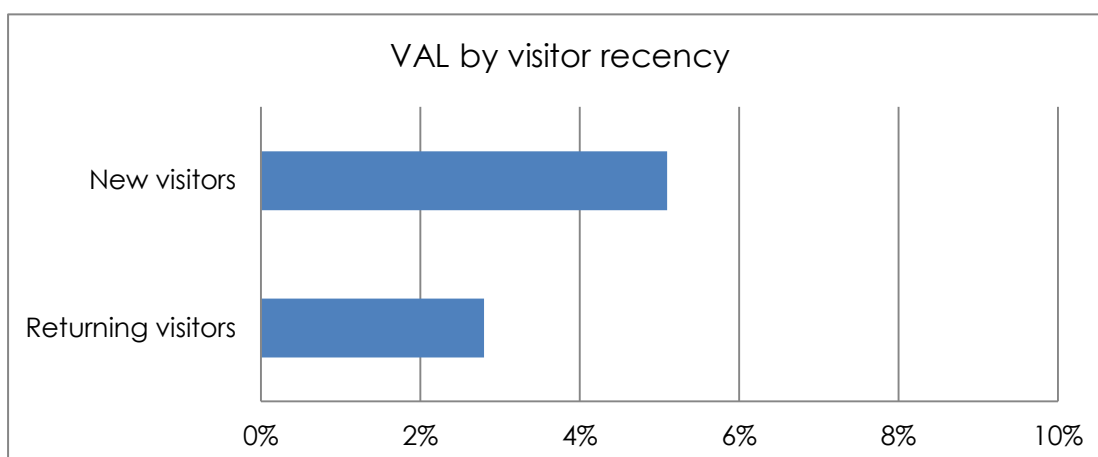
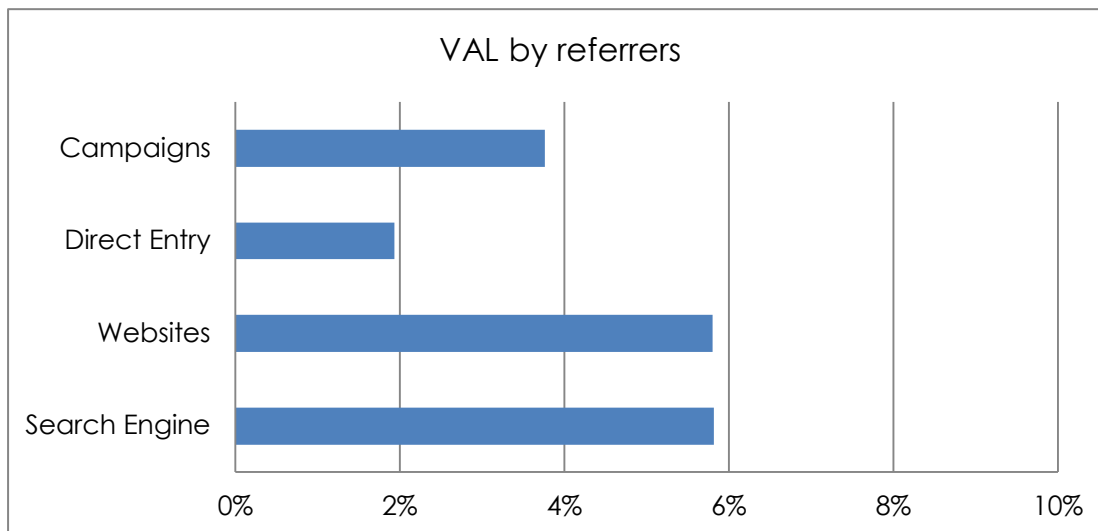
<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=premiere-guerre-mondiale>

<http://www.arch.be/index.php?l=fr&m=actualites&r=toutes-les-actualites&cat=numerisation>

<http://www.arch.be/index.php?l=fr&m=enseignant&r=thematiques-historiques>

<http://www.arch.be/index.php?l=fr&m=chercheur&r=jalons-de-recherche>

Taux de conversion par segment



CAT = catalogue principal = 38,7%

Cela signifie que 38,7% des visites (56 618) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://search.arch.be/fr/rechercher-des-archives>

<http://search.arch.be/fr/producteurs-darchives>

<http://search.arch.be/fr/rechercher-des-personnes>

<http://search.arch.be/fr/tips/101-etat-civil>

<http://search.arch.be/fr/tips/98-registresparoissiaux>

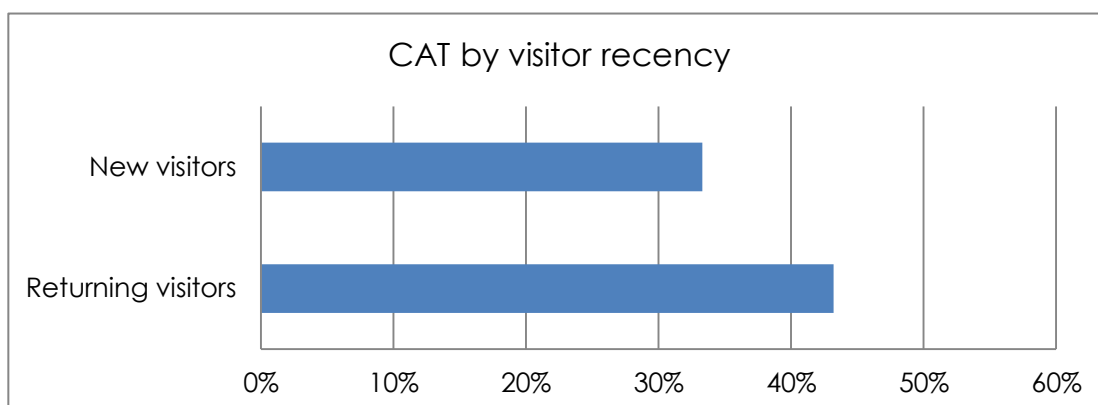
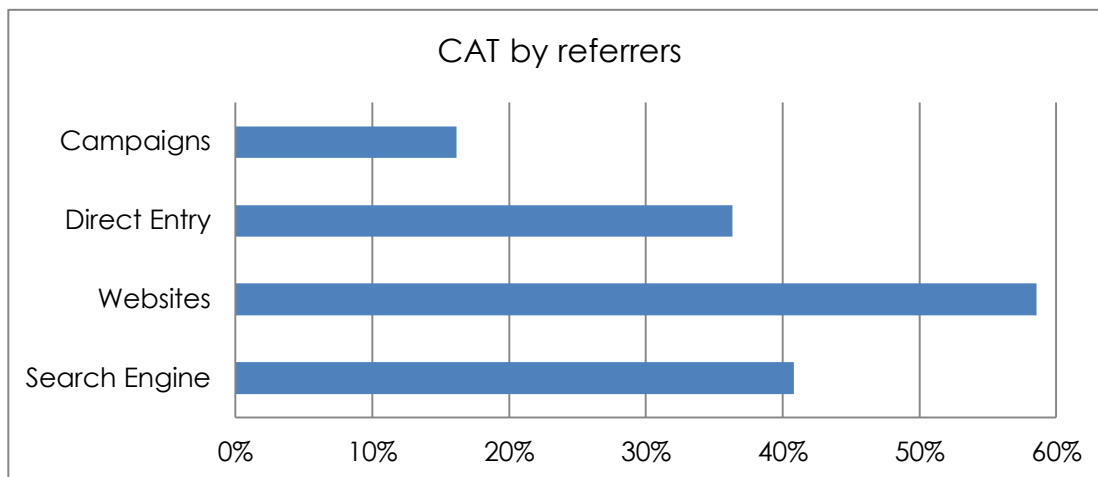
<http://search.arch.be/fr>

<http://search.arch.be/fr/component/user/register>

*Etc.*²⁸

²⁸ Toutes les URLs composées de search.arch.be sont prises en compte grâce à l'utilisation d'expressions régulières.

Taux de conversion par segment



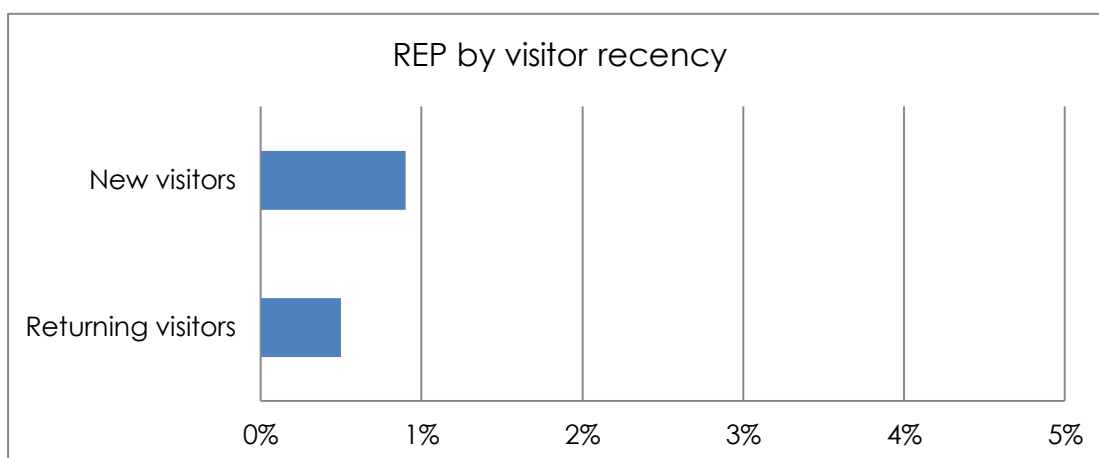
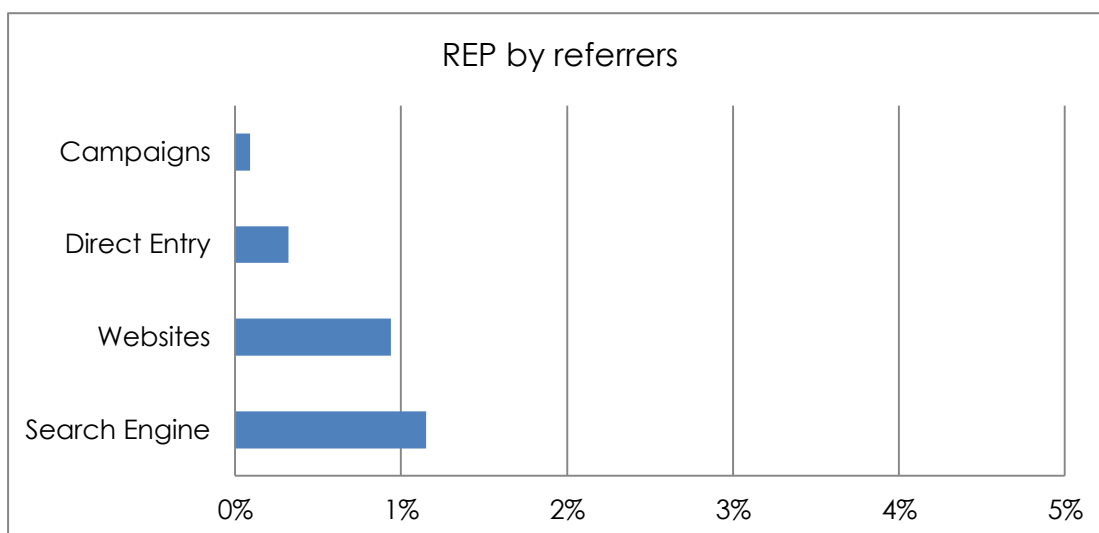
REP = commande de reproductions = 0,6%

Cela signifie que 0,6% des visites (940) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=en-pratique&r=reproductions>

<http://www.arch.be/index.php?l=fr&m=enseignant&r=reproductions>

Taux de conversion par segment



PROFILES = Informations par profil de visiteurs = 28,3%

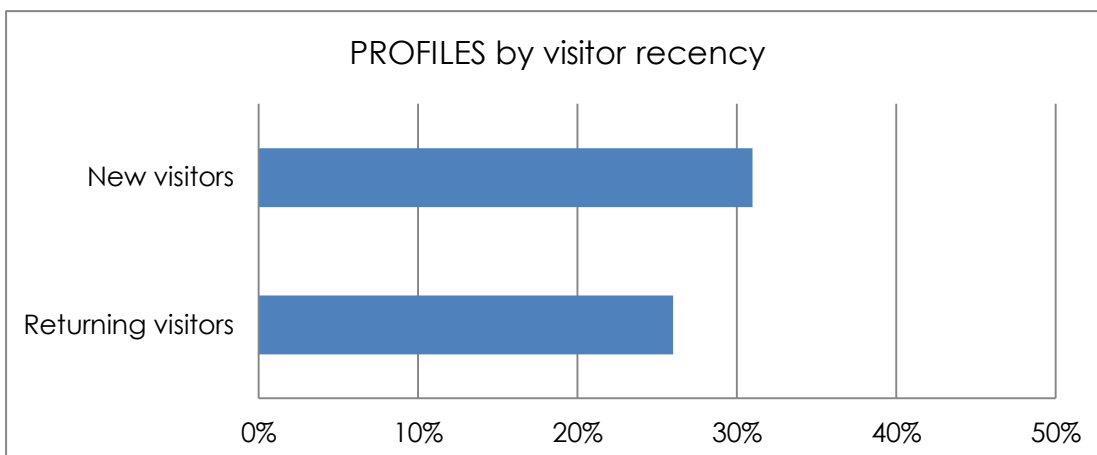
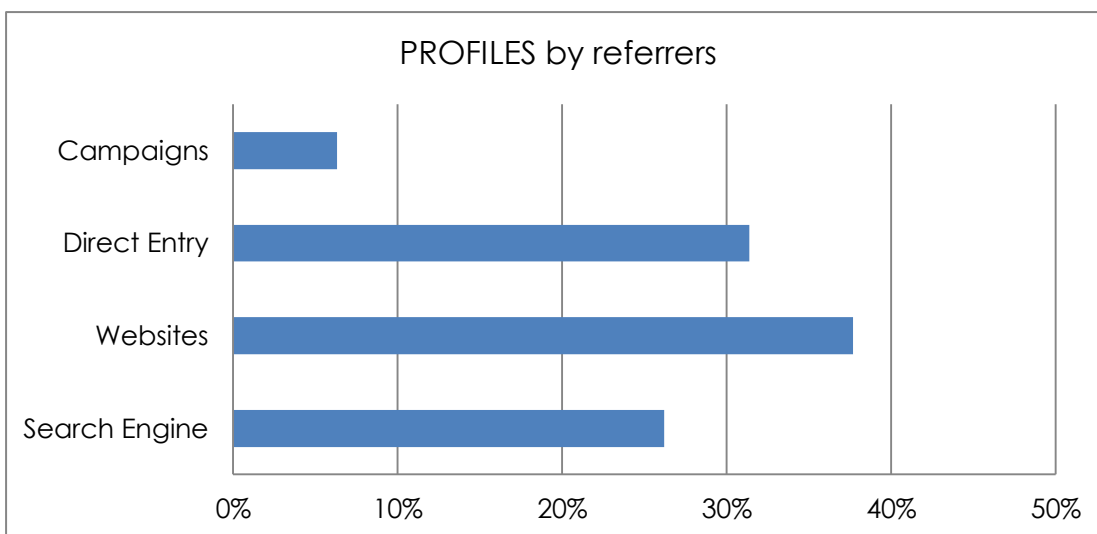
Cela signifie que 28,3% des visites (41 331) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

Toutes les pages concernant un profil en particulier.

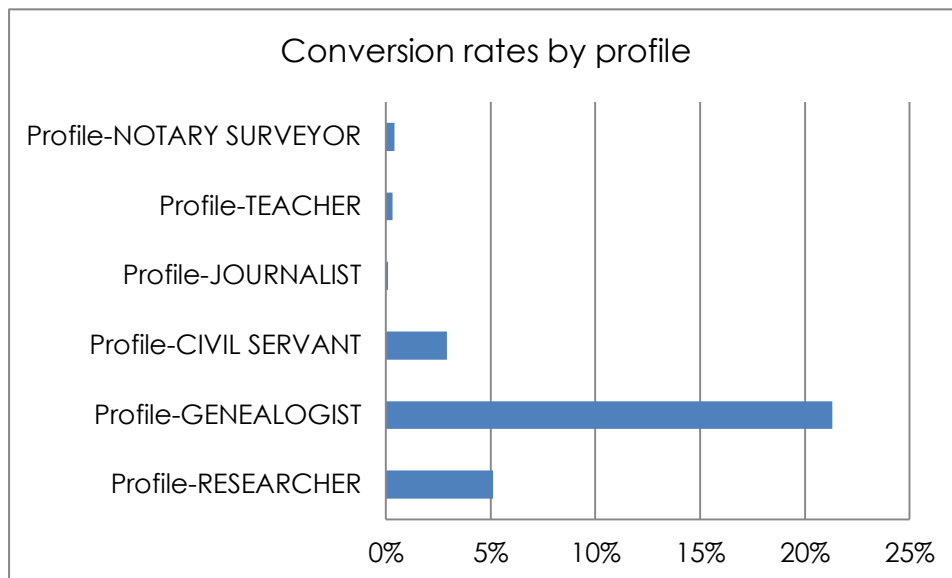
Voici l'expression régulière utilisée :

*. *arch.be\|. * & m =. *(chercheur|researcher|onderzoeker|forscher|genealogist.|genealoog|genealoge|fonctionnaire|ambtenaar|beamte|civil-servant|journalist|enseignant|lehrkraft|leerkracht|teacher|notaire-geometre|notar(is-landmeter|-landmesser|y-surveyor)). **

Taux de conversion par segment



Taux de conversion pour chaque profil (moyenne)



TRI = record schedules = 1,7%

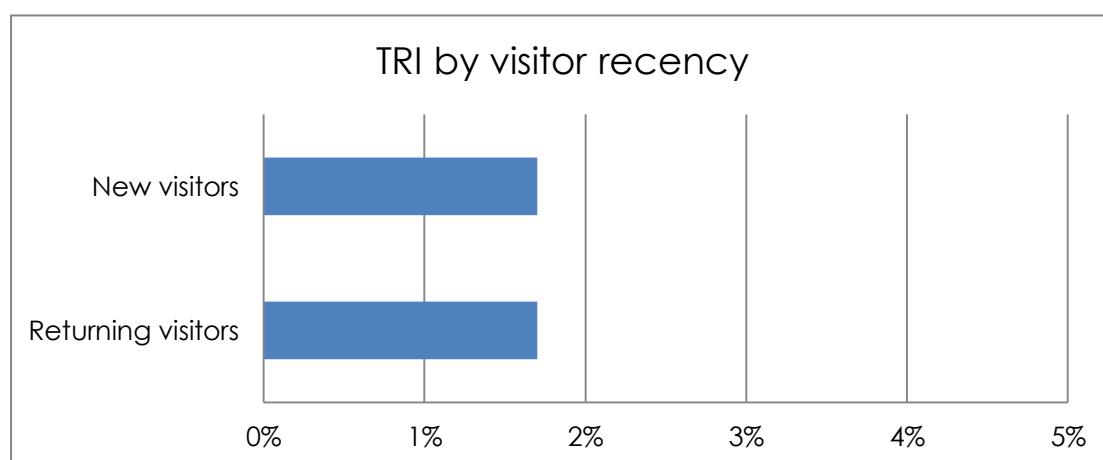
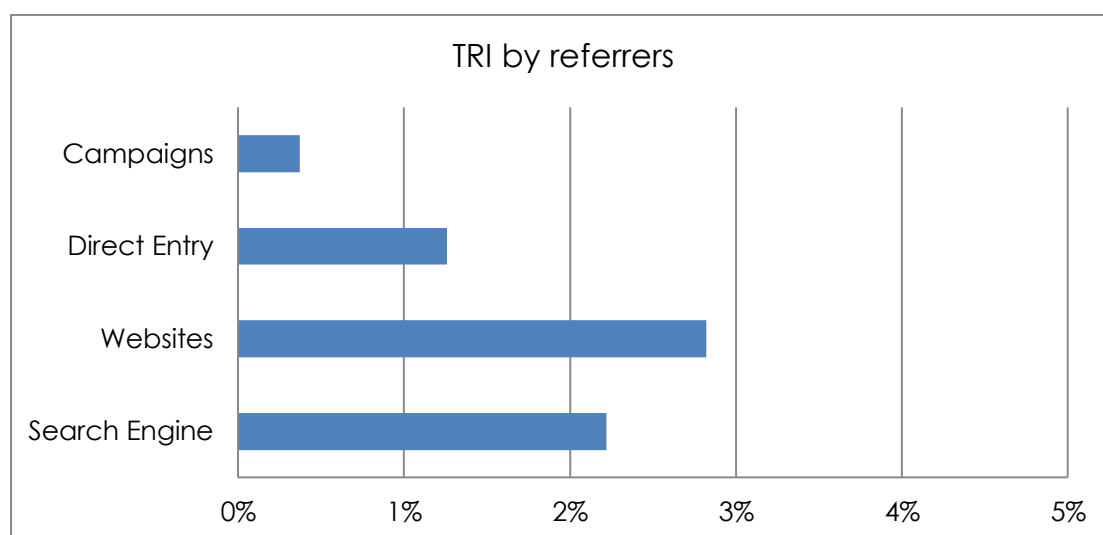
Cela signifie que 1,7% des visites (2 459) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais, anglais ou allemand) :

<http://www.arch.be/index.php?l=fr&m=ressources-en-ligne&r=tableaux-de-tri>

<http://www.arch.be/index.php?l=fr&m=fonctionnaire&r=trier-eliminer-des-documents&sr=tableaux-de-tri&p=benelux>

Etc.²⁹

Taux de conversion par segment



²⁹ L'utilisation d'expressions régulières permet d'englober toutes les URLs similaires à celles-ci.

1.2.2. Analyse spécifique : www.kbr.be



COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 135 000 reprises.

Il est important de noter que parmi ces 135 000 visites, près de la moitié se sont terminées après que la visiteur n'a vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 49% pour le site web de la KBR.

Par ailleurs, il nous est possible de savoir qu'un peu moins de 40% des visiteurs sont venus plus d'une fois sur le site (38,0%), près d'un visiteur sur cinq est venu quatre fois ou plus (19,7%) et 3,1% des visiteurs sont venus plus de 50 fois.



QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que près de quatre visites sur cinq (79,6%) sont effectuées depuis la Belgique, qui est suivie par la France (5,2%) et les Pays-Bas (3,3%), pour ne citer que les pays les plus représentés. Parmi les visites provenant de Belgique, nous pouvons obtenir un détail par région³⁰ ; voici les cinq régions venant en tête : Bruxelles-Capitale (19,2%), le Brabant flamand (6,2%), Anvers (4,8%), la Flandre orientale (4%) et Liège (4,2%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : près de la moitié des visites sont effectuées via des navigateurs configurés en français (44,8%), suivis par le néerlandais (32,5%), viennent ensuite l'anglais (14,8%) et l'allemand (2,0%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (84,3%). Parmi les autres modes de connexion, nous savons que 8,8% des visites sont réalisées à partir d'une tablette smartphone et 5,9% via une tablette.

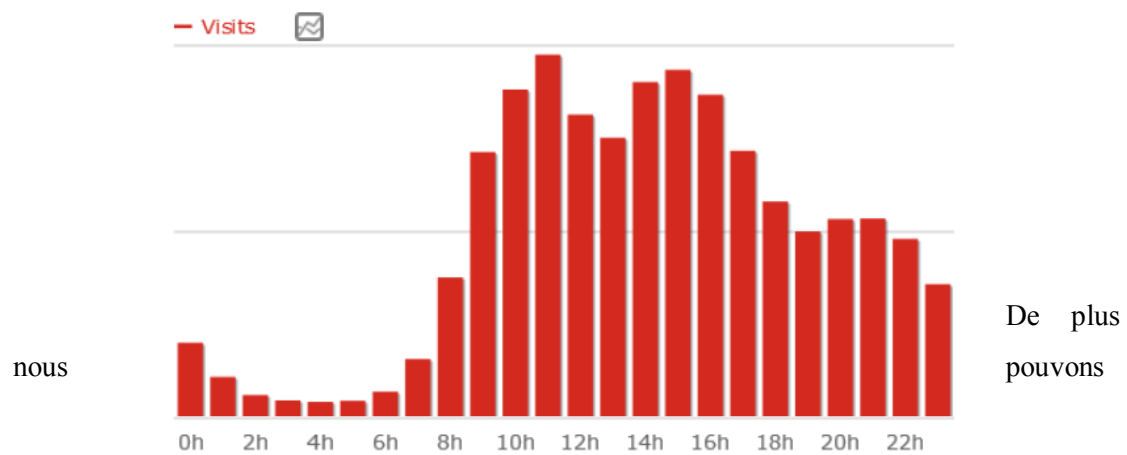
QUELLES SONT LEURS HABITUDES DE VISITE ?

Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (site intéressant qui a su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information recherchée, clics

³⁰ Il faut garder en tête que pour plus de 40% des visites, Piwik n'a techniquement pas été en mesure d'associer l'adresse IP à une région en particulier.

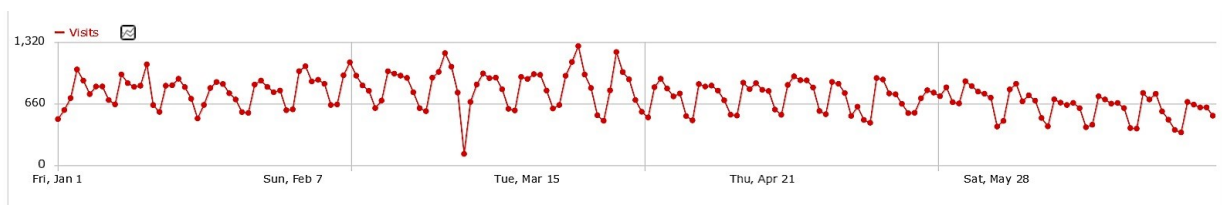
hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu). Parmi toutes les visites effectuées sur le site kbr.be, moins de la moitié (45,9%) durent plus de 10 secondes. Un tiers des visites durent plus d'une minute, 9,2% durent plus de 10 minutes et enfin 2,1% durent plus de 30 minutes. En ce qui concerne le nombre de pages vues, un peu plus de 30% des visites comptabilisent trois pages vues ou plus (32,5%), et seuls 4,2% des visites comptent plus de 10 pages vues.

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 18h, ainsi qu'en début de soirée.



www.kbr.be - Fréquentation par heure (selon l'heure du serveur)

également observer l'évolution des visites à travers le temps. Nous voyons ainsi une baisse de fréquentation régulière durant le week-end, ainsi qu'un grand pic d'affluence autour du 23 mars 2016.



www.kbr.be - Évolution des visites, jour par jour

Des éléments contextuels permettent de comprendre cette affluence : les attentats survenus à Bruxelles le 22 mars et les mesures exceptionnelles que cela a provoqué, notamment la fermeture de différents espaces culturels bruxellois. De nombreuses personnes se sont



probablement connectées sur le site le lendemain des attentats pour savoir si l'établissement était ouvert ou fermé ce jour-là. L'absence d'URLs détaillées³¹ ne nous permet pas de vérifier cette hypothèse. En revanche, la hausse du taux d'accès direct au site et la diminution du nombre de nouveaux visiteurs ce jour-là³² semblent aller dans le sens d'une consultation ciblée, par des utilisateurs préexistants.

À première vue, l'impact de l'envoi de newsletters par l'institution n'est pas flagrant, mais la question nécessiterait d'être creusée.

D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour la Bibliothèque royale, nous savons que 47,7% des visites se font en accès direct, c'est-à-dire que la personne a soit enregistré le site dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Plus de 40% des visites (43,5%) sont le fruit d'une requête lancée sur un moteur de recherche externe. Il est clair que connaître les mots-clés des requêtes ayant conduit au site kbr.be pourrait être très instructif. Malheureusement, la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Nous avons ainsi plus de 55 000 mots-clés « not defined ». Les mots-clés disponibles ayant occasionné plus de 50 visites sont : « bibliothèque royale » (145 visites) ; « bibliothèque royale de belgique » (105 visites) ; « kbr » (94 visites) ; « ferrariskaart » (87 visites) ; « carte ferraris » (67 visites) et « ferraris » (54 visites).

Un peu moins de 10% des visites (8,8%) correspondent à un clic sur un lien vers le site de la Bibliothèque royale effectué sur d'autres sites web/réseaux sociaux. Nous retrouvons

Facebook en tête de liste, cependant le détail est un peu confus et il n'est pas certain que tous les postes Facebook aient été pris en compte.

³¹ Par exemple une URL qui mentionnerait le mot « attentat », « alerte » ou « fermeture ».

³² Le taux de visites issues d'un accès direct au site, qui s'élève en moyenne à 47,7%, atteint ce jour-là les 60%, et le taux de nouveaux visiteurs est seulement de 53% versus une moyenne de 62%.



Enfin, comme expliquée précédemment, Piwik ne reconnaît pas les newsletters envoyée par la KBR. Ces dernières devraient être paramétrées d'une certaine façon pour que Piwik puisse les considérer comme telles.

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE / SORTIE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ces forces et ces faiblesses.

En ce qui concerne les pages d'entrées, les pages issues de la rubrique « collections » sont en tête. Si l'on regarde en détail, c'est la sous-rubrique « cartes et plans » et plus précisément les pages concernant la carte Ferraris qui génèrent beaucoup de trafic. Il apparaît que les trois pages donnant accès à la carte Ferraris (une page par langue : anglais, français et néerlandais) représentent donc la porte d'entrée sur le site pour près de 2 visites sur 10³³ !

Nous retrouvons ces mêmes pages consacrées à la carte Ferraris parmi les pages ayant un taux de sortie élevé.

³³ Ces trois pages correspondent à 23 825 entrées sur le site pour une période durant laquelle 137 244 visites ont été comptabilisées, cela correspond donc à 17,35% des visites.

GOALS

Pour le site kbr.be, 10 goals ont été identifiés :

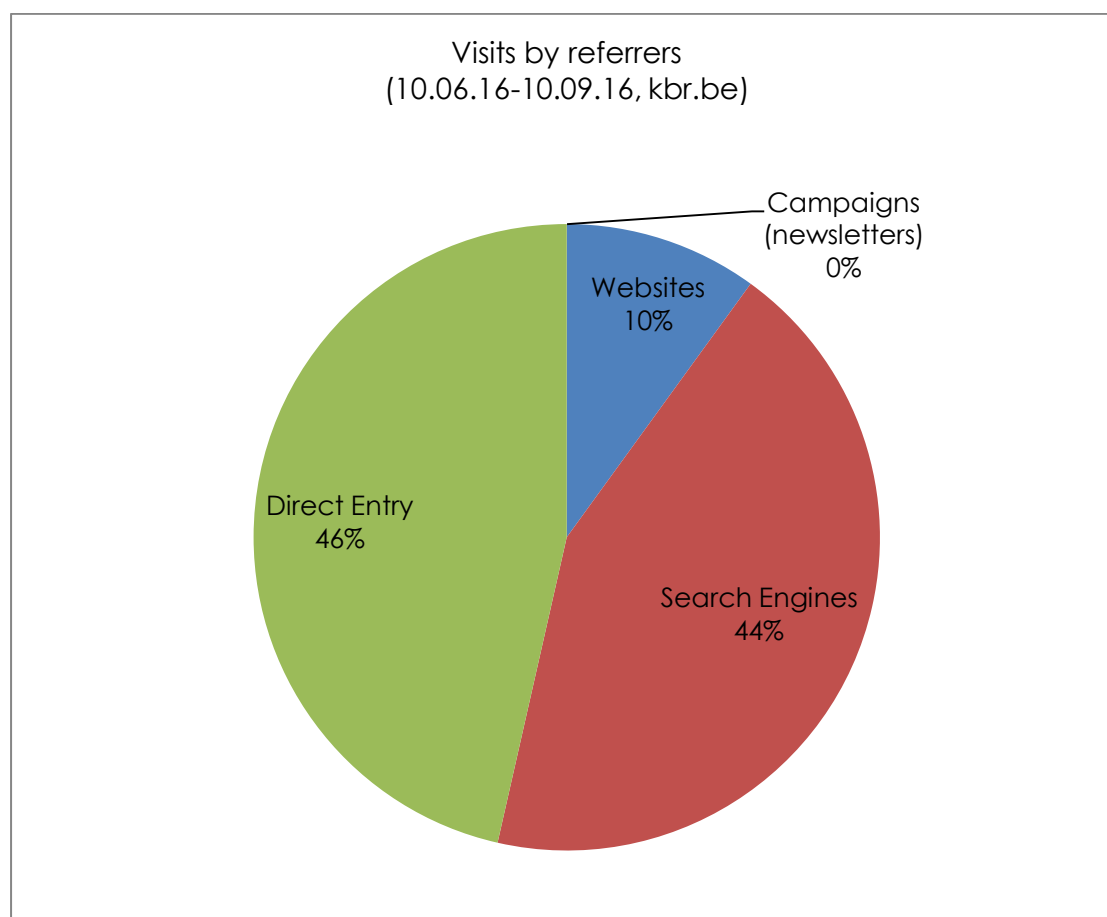
| KBR.BE | | | |
|-------------|-------------------------------------|---|----------------------------------|
| Name | Access to | Web pages that contain | Taux de conversion ³⁴ |
| PRA | Practical information | Contact info, address, opening hours, prices of reading room cards | 3,4% |
| CONT I | Digital/digitised heritage content | Items of heritage collections | 17,4% |
| CONT II | Meta-information | Historical background, institution's mission and structure | 41,9% |
| CONT II-PUB | Publications | Information related to the KBR publications | 0,8% |
| CONT III | News and agenda | Agenda, news, newsletter subscription | 9,0% |
| TUT | Help and tutorial | FAQ, tutorials, site and search engine manuals, help pages, search assistance | 3,4% |
| TUT-PDF | Help and tutorial | FAQ, tutorials, site and search engine manuals, help pages, search assistance (PDF) | 0,5% |
| CAT | Main catalogue | Hyperlinks to main digital catalogue | (5,5%) |
| MET-CAT | Metadata related to the collections | Metadata related to the collections (inside main digital catalogue) | 2,1% |
| REP | Ordering reproductions | Information related to ordering reproductions | 2,7% |

Les taux de conversion présentés dans les pages qui suivent proviennent tous d'une période d'analyse de 3 mois (10.06.2016-10.09.2016)³⁵.

³⁴ Moyenne.

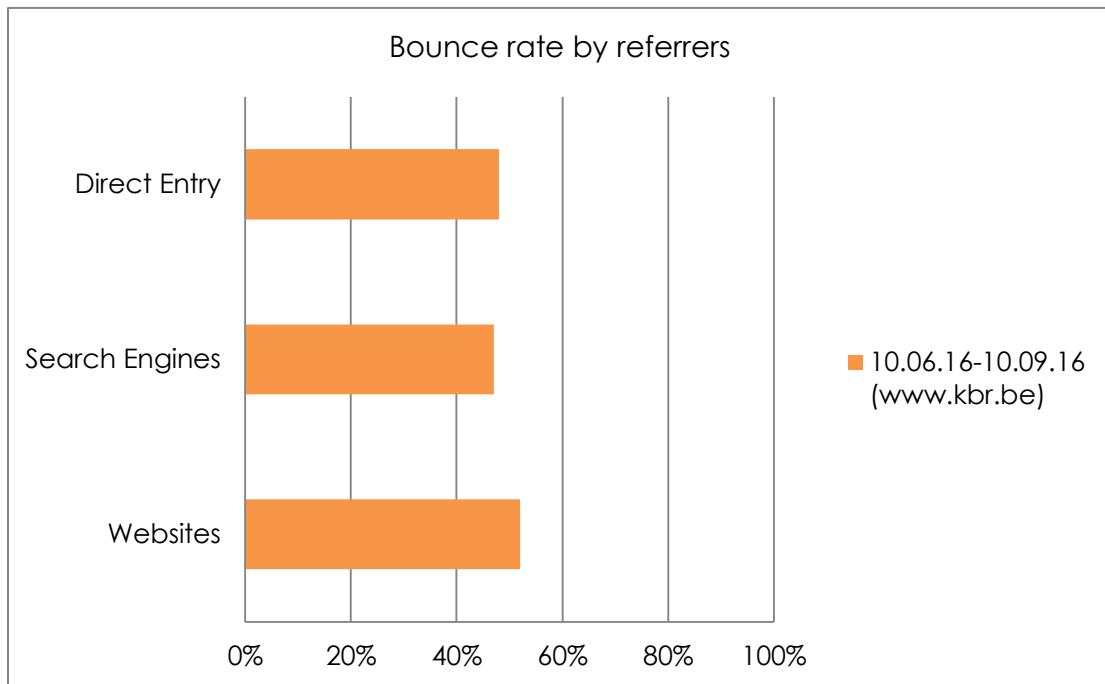
Durant ce laps de temps, 54 267 visites ont été comptabilisées. Avant de présenter les taux de conversion par goal, il semble important de d'abord se pencher sur l'origine de ces visites.

Segmentation des visites effectuées entre le 10.06.2016 et le 10.09.2016 :



Par ailleurs, si l'on veut se pencher sur les taux de conversion d'un site, il est utile de garder en tête le bounce rate de chaque segment (si l'on sait par exemple que plus de la moitié des utilisateurs quittent le site dès la page d'accueil, alors un taux de conversion de 20% semble directement plus encourageant).

³⁵ Les goals, qui ne sont pas rétroactifs ont été mis en place au début du mois de juin 2016, au terme d'un long processus préparatoire. C'est pour cette raison que la période d'analyse n'est pas la même que pour la première partie du rapport.





PRA = information pratique = 3,4%

Cela signifie que 3,4% des visites (1 829) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

http://www.kbr.be/informations/informations_fr.html

http://www.kbr.be/contacts_fr.html

http://www.kbr.be/informations/heures_fr.html

http://www.kbr.be/services/photographie/photo_contact_fr.html

http://www.kbr.be/services/photographie/photo_contact_en.html

http://www.kbr.be/collections/depot_legal/directives_fr.html

http://www.kbr.be/collections/depot_legal/contacts_fr.html

http://www.kbr.be/collections/reserve/reserve_consult_fr.html

http://www.kbr.be/collections/estampes/informations/localisation_fr.html

<http://www.kbr.be/educ/visitesGuidees.html>

http://www.kbr.be/collections/cart_plan/informations/informations_fr.html

http://www.kbr.be/collections/imprimésDrukken/contact_fr.html

http://www.kbr.be/collections/journaux/informations/contacts_fr.html

http://www.kbr.be/collections/reserve/reserve_contact_fr.php

http://www.kbr.be/collections/musique/informations/infoContacts_fr.html

http://www.kbr.be/collections/chalco/chalco_contact_fr.html

http://www.kbr.be/collections/medailles/informations/contacts_fr.html

http://www.kbr.be/collections/musique/informations/infoContacts_fr.html

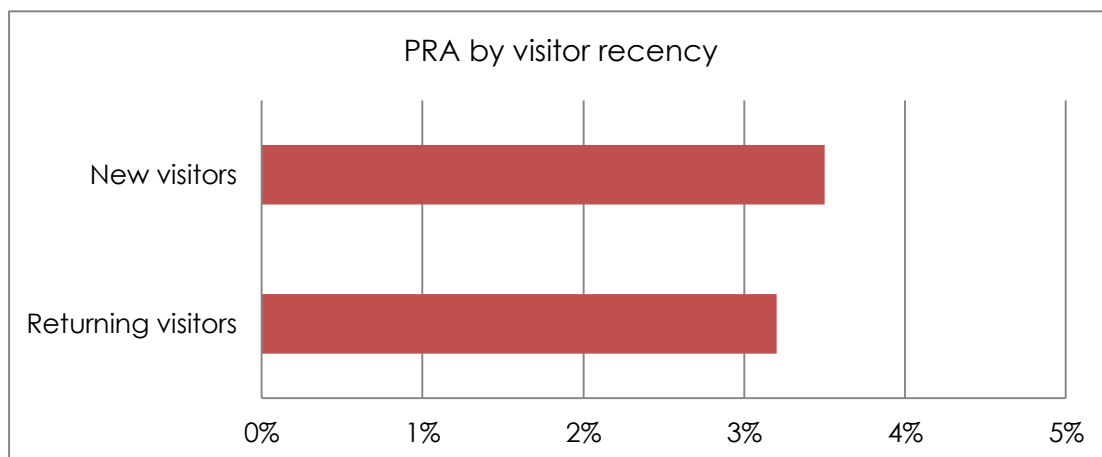
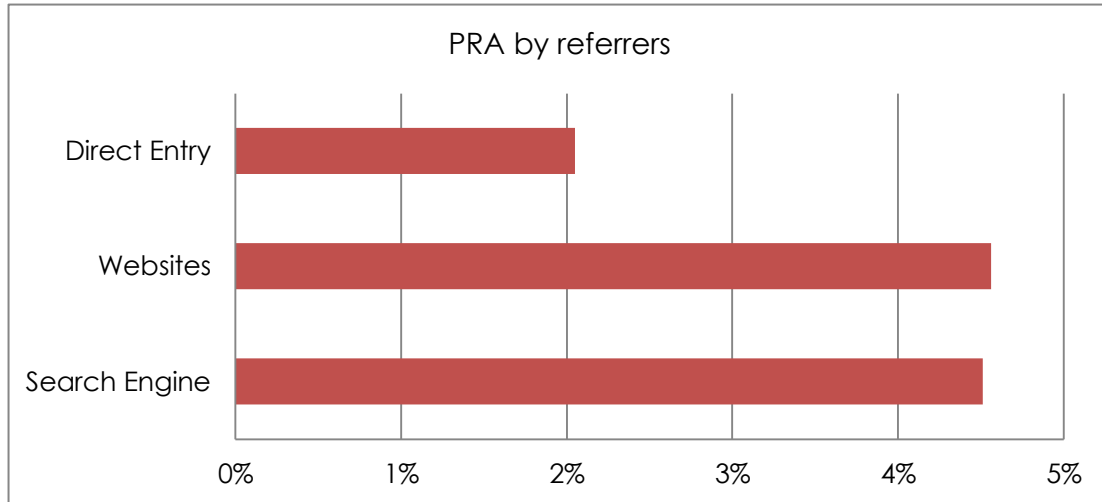
http://www.kbr.be/collections/estampes/informations/localisation_fr.html

http://www.kbr.be/collections/chalco/chalco_contact_fr.html

http://www.kbr.be/collections/medailles/informations/contacts_fr.html



Taux de conversion par segment

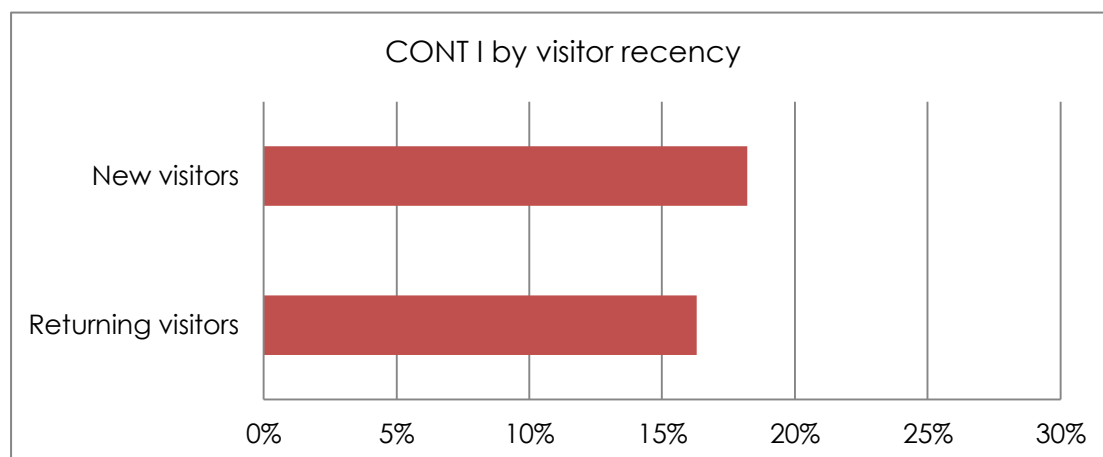
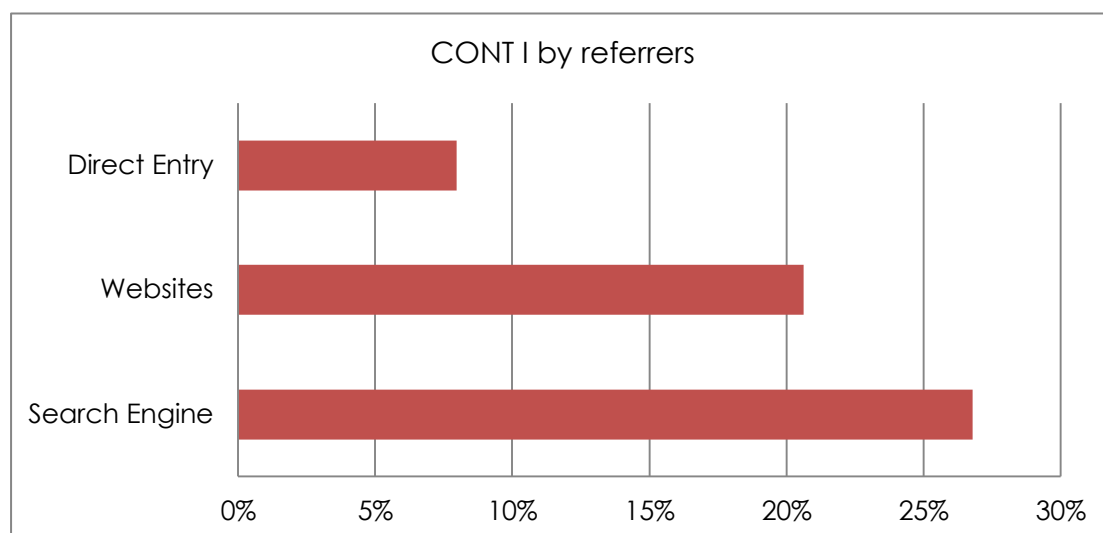


CONT I = collections numérisées = 17,4%

Cela signifie que 17,4% des visites (9 458) comportent au moins un passage sur l'une de ces pages (ou sur la page équivalente en néerlandais) :

http://www.kbr.be/collections/cart_plan/ferraris/ferraris_fr.html

Taux de conversion par segment



CONT II = Meta-information = 41,9%

Cela signifie que 41,9% des visites (22 752) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

http://www.kbr.be/informations/histoire_fr.html

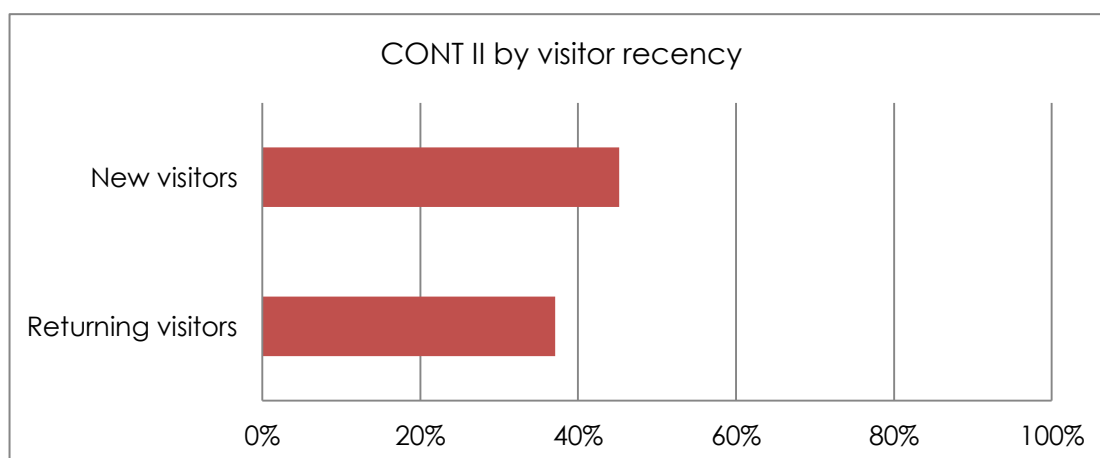
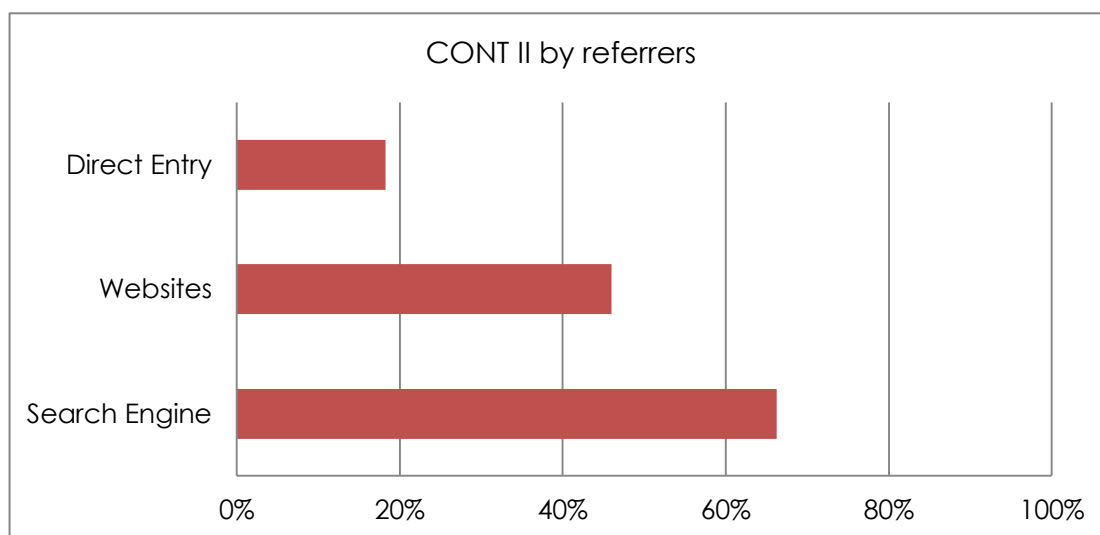
http://www.kbr.be/informations/missions/missions_fr.html

http://www.kbr.be/collections/cart_plan/collections/collections_fr.html

http://www.kbr.be/collections/cart_plan/cart_plan_fr.html

Etc.³⁶

Taux de conversion par segment



³⁶ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires (ici, il s'agit des différentes sous-rubriques de « collections »).



CONT II-publications = Publications = 0,8%

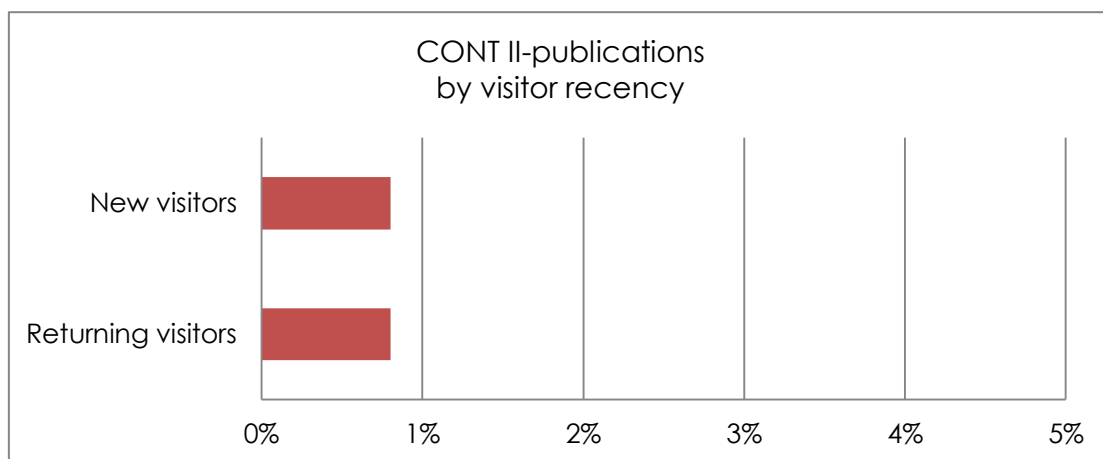
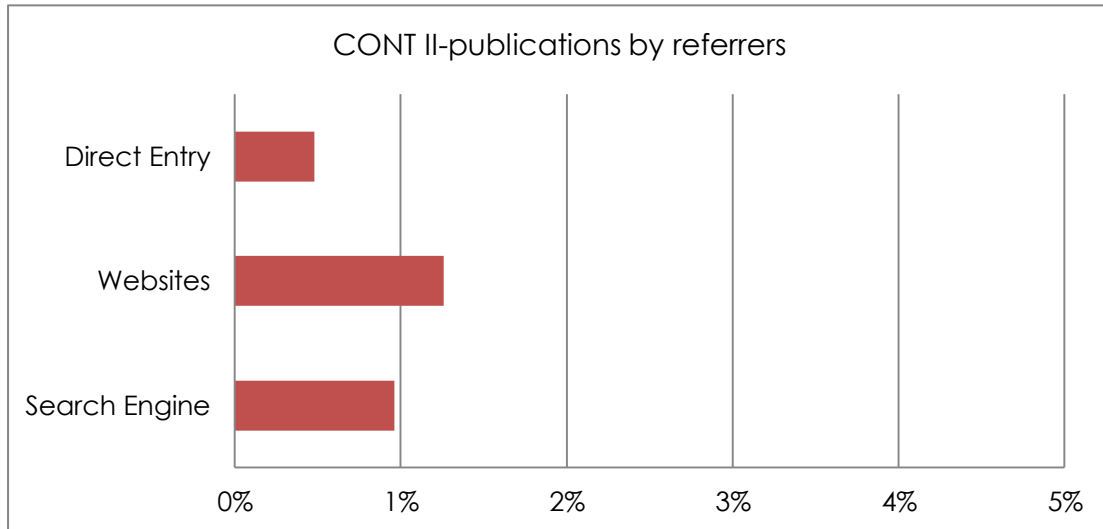
Cela signifie que 0,8% des visites (416) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

http://www.kbr.be/informations/publications/inMonteArtium_fr.html
http://www.kbr.be/informations/publications/monografieen_fr.html
http://www.kbr.be/informations/publications/cat_expo_fr.html
http://www.kbr.be/informations/publications/cas_fr.html
http://www.kbr.be/informations/publications/dossier_fr.html
http://www.kbr.be/informations/publications/speciales_fr.html
http://www.kbr.be/informations/publications/rapportAnnuel/rapportAnnuel_fr.html
http://www.kbr.be/informations/publications/inMonteArtium_fr.html
http://www.kbr.be/informations/publications/monografieen_fr.html
http://www.kbr.be/informations/publications/cat_expo_fr.html
http://www.kbr.be/informations/publications/cas_fr.html
http://www.kbr.be/informations/publications/dossier_fr.html
http://www.kbr.be/informations/publications/speciales_fr.html
*Etc.*³⁷

³⁷ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires (ici, il s'agit des différentes URLs composées de ../informations/publications/...).



Taux de conversion par segment





CONT III = Nouvelles et agenda = 9,0%

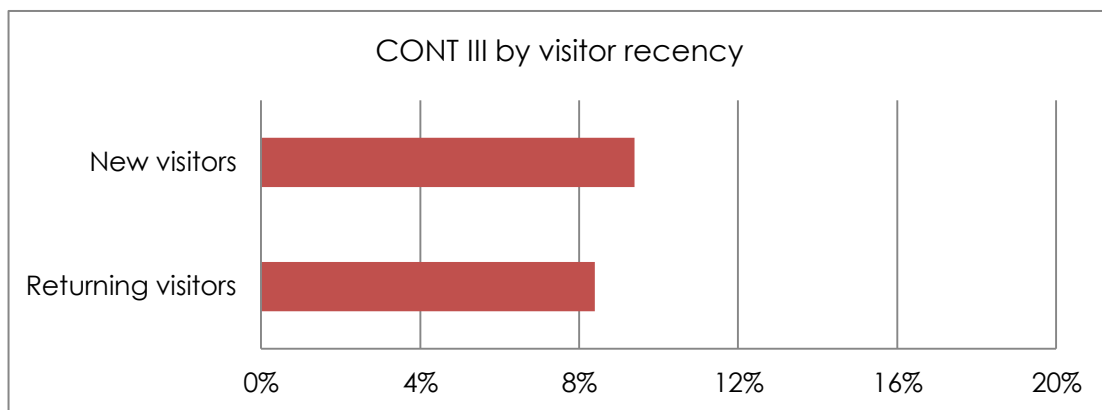
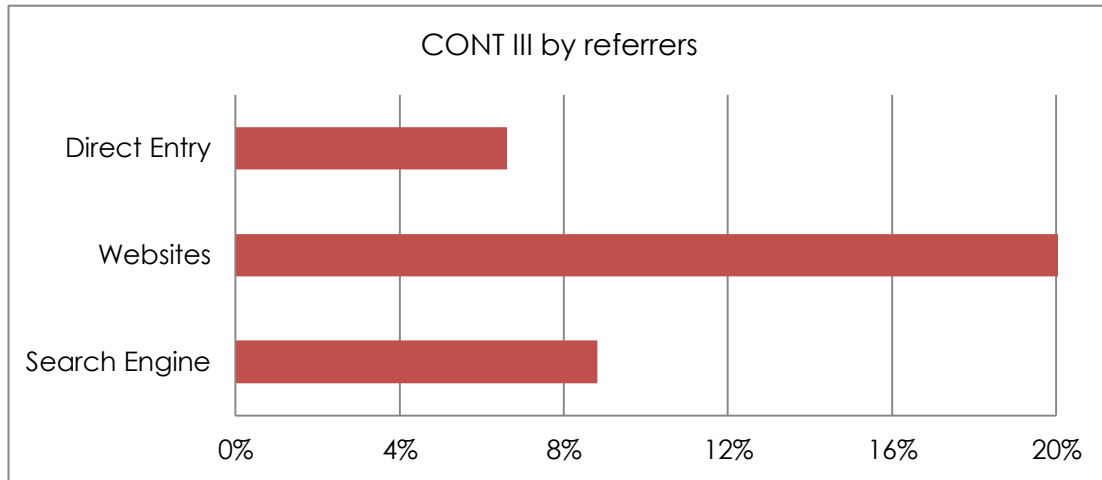
Cela signifie que 9,0% des visites (4 881) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

<http://www.kbr.be/actualites/colloque/bbc/en.php>
<http://www.kbr.be/actualites/10vertissezvous/10vertissezvous.html>
http://www.kbr.be/actualites/concerts/concerts_fr.html
<http://www.kbr.be/actualites/nocturneMusees/fr.html>
http://www.kbr.be/actualites/palace/palace_fr.html
http://www.kbr.be/informations/newsletter/inscription_fr.php
<http://www.kbr.be/actualites/10vertissezvous/10vertissezvous.html>
http://www.kbr.be/actualites/concerts/concerts_fr.html
*Etc.*³⁸

³⁸ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires (ici, il s'agit des différentes URLs débutant par [kbr.be/actualites/...](http://www.kbr.be/actualites/))



Taux de conversion par segment





TUT = Aide et tutoriels = 3,4%

Cela signifie que 3,4% des visites (1 829) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

http://www.kbr.be/collections/depot_legal/depot_legal_fr.html

http://www.kbr.be/catalogues/Edepot/edepot_fr.html

http://www.kbr.be/services/depot_legal/depot_legal_fr.html

http://www.kbr.be/informations/publications/guideGids/guide_fr.html

http://www.kbr.be/collections/cart_plan/salle_lecture/salle_lecture_fr.html

http://www.kbr.be/collections/cart_plan/outils_recherche/outils_recherche_fr.html#princip

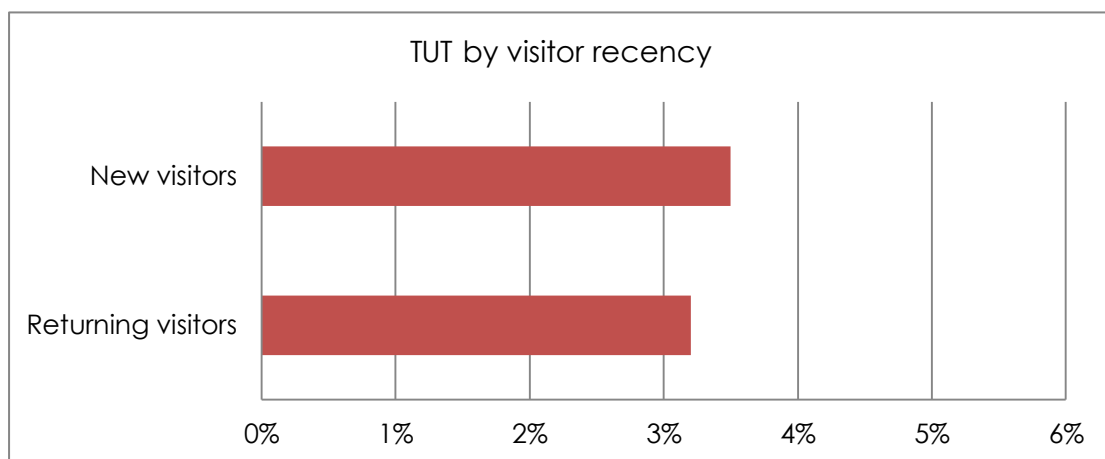
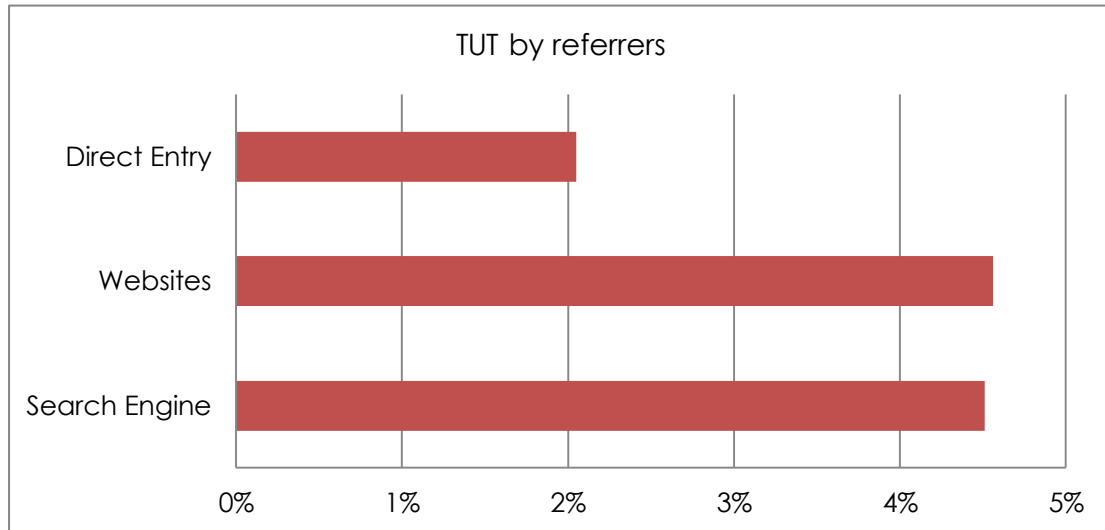
http://www.kbr.be/collections/cart_plan/salle_lecture/salle_lecture_fr.html

http://www.kbr.be/collections/depot_legal/depot_legal_fr.html

http://www.kbr.be/catalogues/Edepot/edepot_fr.html



Taux de conversion par segment



TUT-PDF³⁹ = Aide et tutoriels pdf = 0,5%

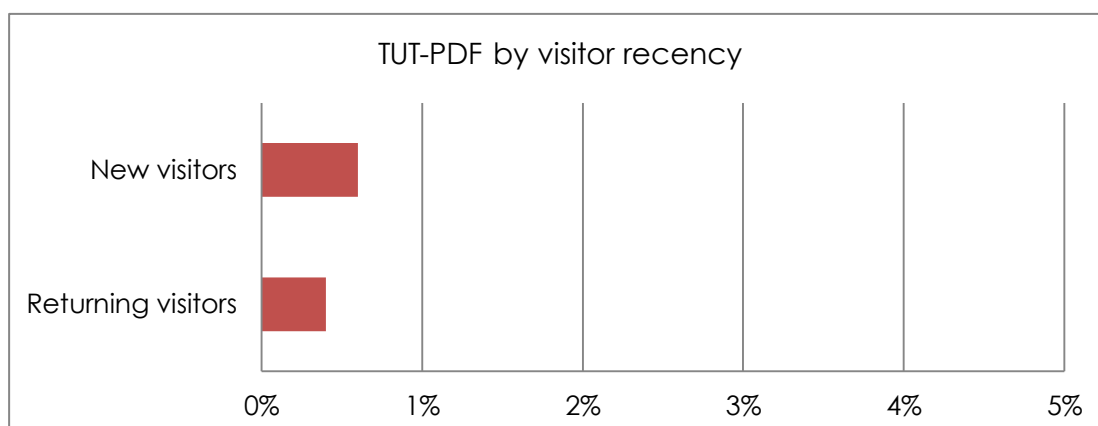
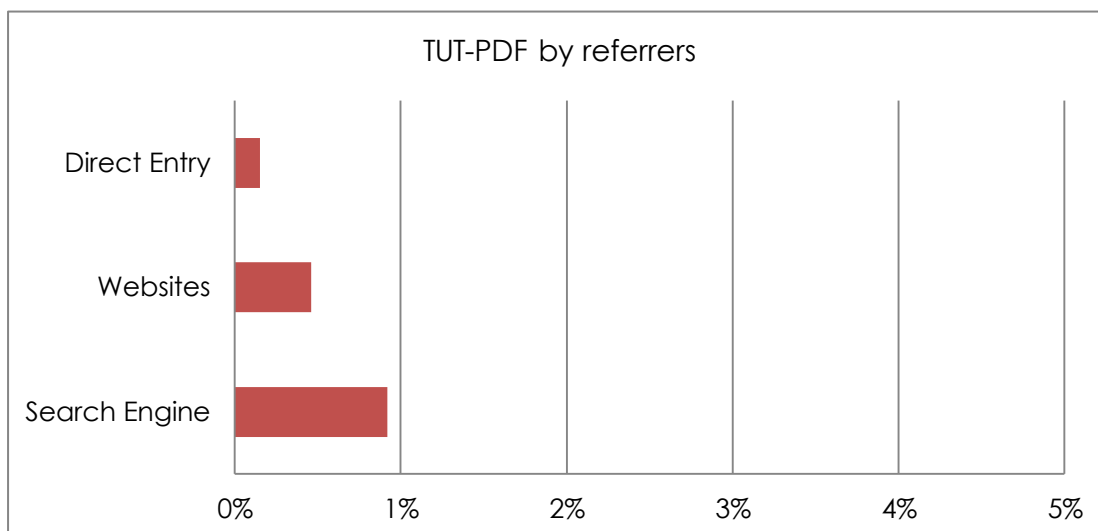
Cela signifie que 0,5% des visites (281) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

<http://belgica.kbr.be/pdf/jour/manuelUtilisateur.pdf>

http://www.kbr.be/collections/cart_plan/salle_lecture/consult_enligne2.ppt

http://www.kbr.be/collections/cart_plan/salle_lecture/consult_enligne1.ppt

Taux de conversion par segment



³⁹ Ici il s'agit également de fichiers PPT.



CAT = catalogue principal = 5,5%

Cela signifie que 5,5% des visites (2 966) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais) :

<http://www.opac.kbr.be>

+ toutes les autres URLs basées sur cette URL, exemple :

<http://opac.kbr.be/belgicapress.php?lang=FR>

Attention : il s'est avéré que ce goal ne couvre qu'une partie des renvois vers le catalogue ! En effet, il est apparu a posteriori que ces URLs ne regroupaient pas toutes les visites vers le catalogue. Il manque toute une série de liens renvoyant encore vers l'ancien opac de la KBR (« opteron »)⁴⁰. Il aurait donc fallu compléter le goal à l'aide d'URLs commençant par *[http://opteron1.kbr.be/...](http://opteron1.kbr.be/)* Ce chiffre de 5,5% n'est donc pas représentatif de la réalité : il devrait être plus élevé. Malheureusement, les goals ne sont pas rétroactifs ; les modifier maintenant ne permettrait donc pas de corriger les chiffres présentés ci-dessus.

Concrètement, c'est la confrontation des chiffres obtenus via ce goal avec d'autres données similaires qui a révélé un écart anormalement élevé. Cette étape de vérification consiste à comparer trois sources d'information :

- le « **taux de conversion** » obtenu grâce à la mise en place d'un goal
- le nombre de « **liens sortants** » du site www.kbr.be qui renvoient vers l'opac KBR
- le nombre d'entrées sur l'opac KBR provenant du site www.kbr.be

Si l'on reprend la période d'analyse utilisée pour les goals, soit du 10.06.16 au 10.09.16, et le nombre total de visites pour cette période (54 267), voici les différentes données que l'on obtient :

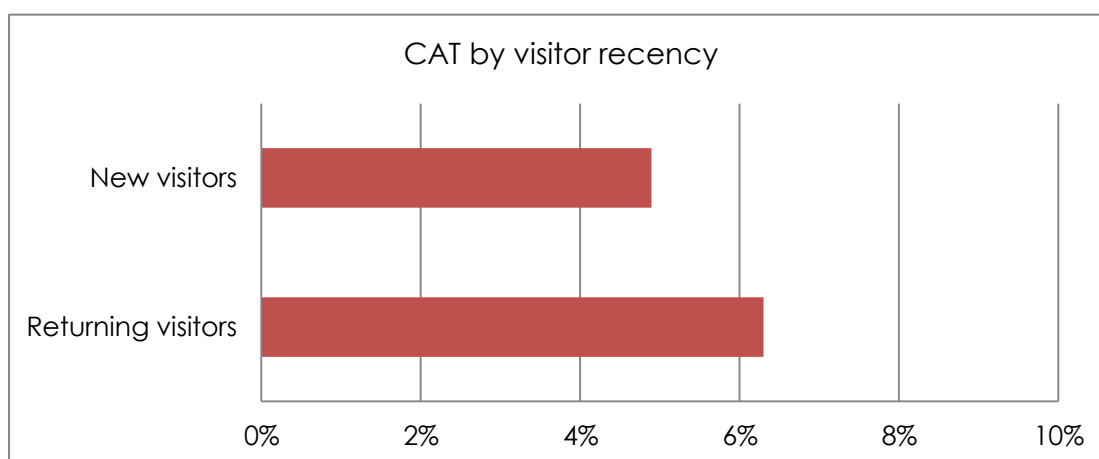
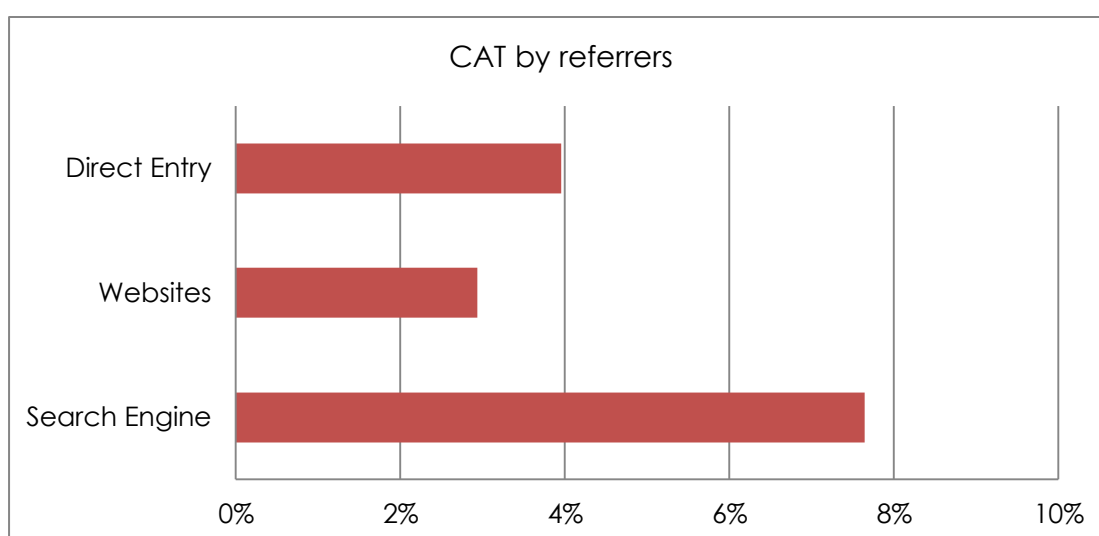
| | Goal CAT | Liens sortants de kbr.be dirigés vers l'opac KBR ⁴¹ | Les entrées sur l'opac KBR issues du site kbr.be |
|---------------|----------|--|--|
| Nombres | 2 966 | 5 132 | 15 136 |
| Total visites | 5,46% | 9,46% | 27,89% |

⁴⁰ Ces liens ont été pris en compte de façon isolée au sein du goal suivant : « MET-CAT » (clics vers des métadonnées figurant au sein du catalogue principal).

⁴¹ On compte chaque clic 1x maximum par visite, à l'instar de ce qui est fait pour les goals.

S'il est fréquent d'avoir de légers écarts lorsque l'on compare des résultats obtenus par des biais différents, y compris des résultats compilés automatiquement par Piwik sans intervention humaine (par exemple les liens entrants d'un site A vers un site B et les liens sortants d'un site B vers un site A)⁴², cette fois-ci l'écart était anormalement élevé et a permis d'identifier la présence de ces URLs renvoyant vers l'ancien opac de la KBR.

Taux de conversion par segment



⁴² À l'heure actuelle, il ne nous est pas possible de comprendre ou d'éviter ce phénomène étant donné qu'avant d'afficher ces résultats, Piwik effectue une agrégation des données brutes, sur laquelle nous n'avons pas vraiment de prise ou de contrôle.

MET-CAT = métadonnées du catalogue principal = 2,1%

Cela signifie que 2,1% des visites (1 125) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais):

[http://opteron1.kbr.be/cgi-](http://opteron1.kbr.be/cgi-bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=6&P3=R_KI&P6=1867882)

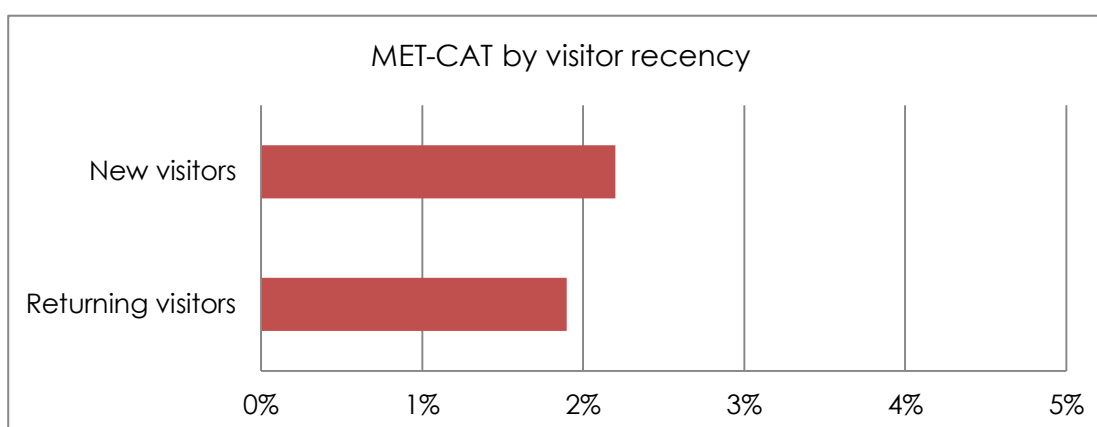
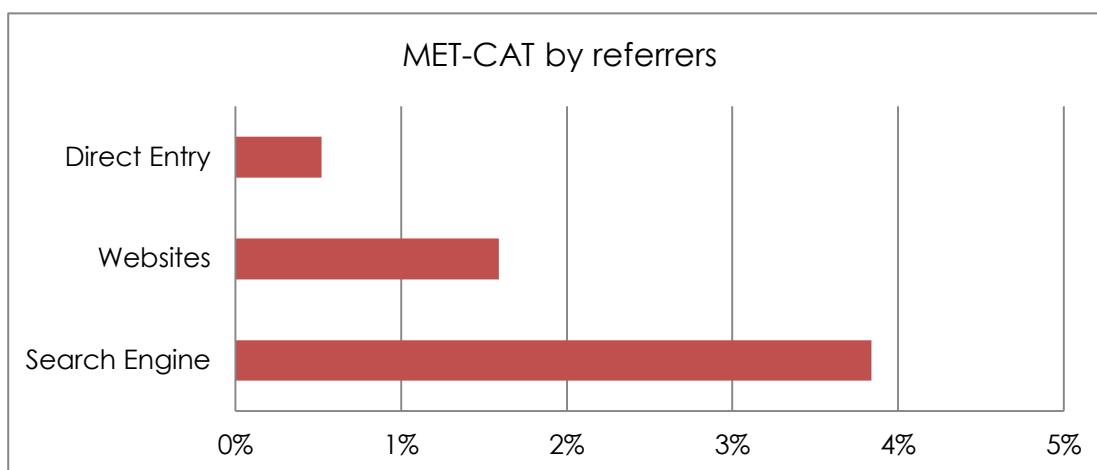
[bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=6&P3=R_KI&P6=1867882](http://opteron1.kbr.be/cgi-bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=6&P3=R_KI&P6=1867882)

[http://opteron1.kbr.be/cgi-](http://opteron1.kbr.be/cgi-bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=5&P3=R_KI&P6=1588624)

[bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=5&P3=R_KI&P6=1588624](http://opteron1.kbr.be/cgi-bin/opac.cgi?P0=FKBR&P1=1_JAN&P9=&P5=20&P4=&P2=5&P3=R_KI&P6=1588624)

Etc.⁴³

Taux de conversion par segment



⁴³ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires (ici, il s'agit des différentes URLs débutant par <http://opteron1.kbr.be/cgi-bin/opac.cgi>....)



REP = Commande de reproductions = 2,7%

Cela signifie que 2,7% des visites (1 461) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.kbr.be/services/chalco_fr.html

http://www.kbr.be/services/prets_expo_fr.html

http://www.kbr.be/services/shop_fr.html

http://www.kbr.be/collections/cart_plan/reprographie/reprographie_fr.html

http://www.kbr.be/services/reserv_pub_fr.html

http://www.kbr.be/services/photographie/photographie_fr.html

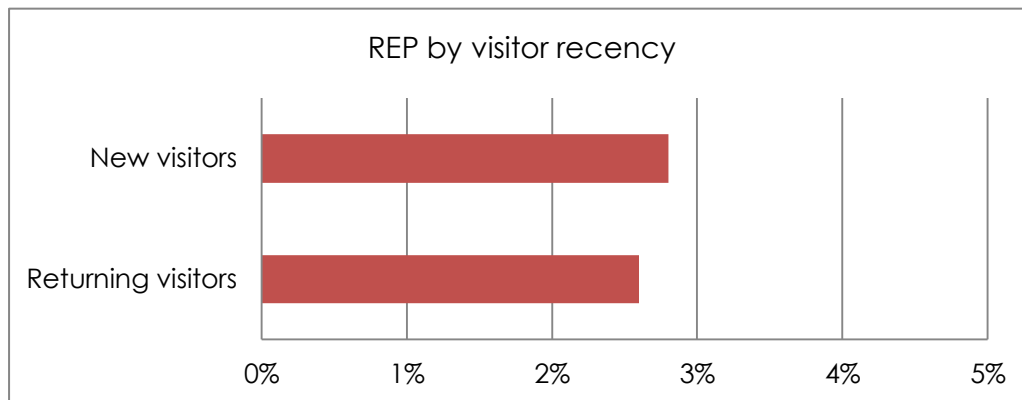
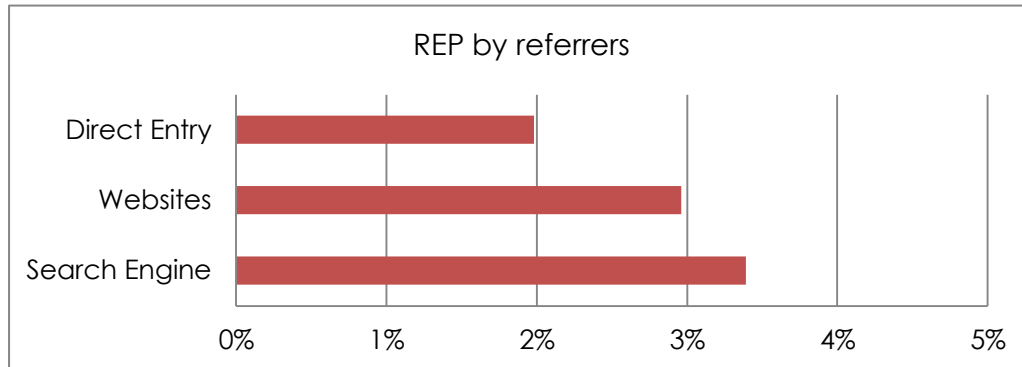
http://www.kbr.be/services/photographie/photo_tarif_fr.html

http://www.kbr.be/services/photographie/photo_regl_fr.html etc.⁴⁴

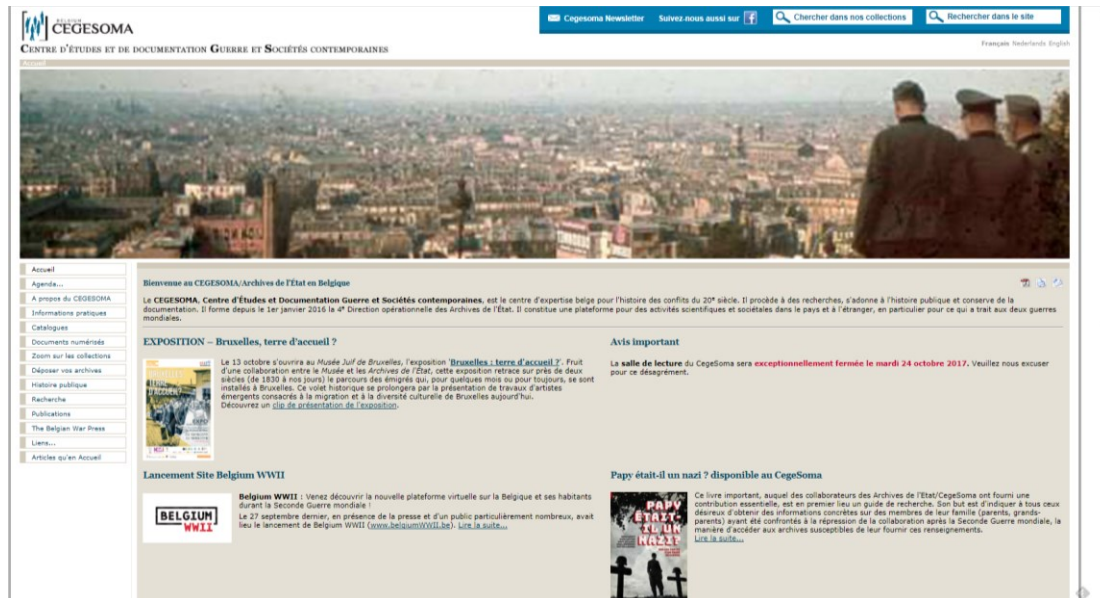
⁴⁴ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires (ici, il s'agit des différentes URLs débutant par [http://www.kbr.be/services/photographie/...](http://www.kbr.be/services/photographie/))



Taux de conversion par segment



1.3.3 Analyse spécifique : www.cegesoma.be



COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 25 000 reprises.

Il est important de noter que parmi ces 25 000 visites, près de la moitié se sont terminées après que la visiteur n'a vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 46% pour le site web du Cegesoma.

Par ailleurs, il nous est possible de savoir que près de 40% des visiteurs sont venus plus d'une fois sur le site (38,9%), plus d'un visiteur sur cinq est venu quatre fois ou plus (23,3%) et 5% des visiteurs sont venus plus de 50 fois.

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que plus de trois visites sur quatre (75,8%) sont effectuées depuis la Belgique, qui est suivie par la France (6,7%) et les Pays-Bas (3,1%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 40% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (11,5%), Anvers (7,4%), le Brabant flamand (6,8%), la Flandre orientale (5,9%) et Liège (4,2%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la

langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (38,6%), suivis de près par le néerlandais (34,8%), viennent ensuite l'anglais (18,7%) et l'allemand (3,6%).

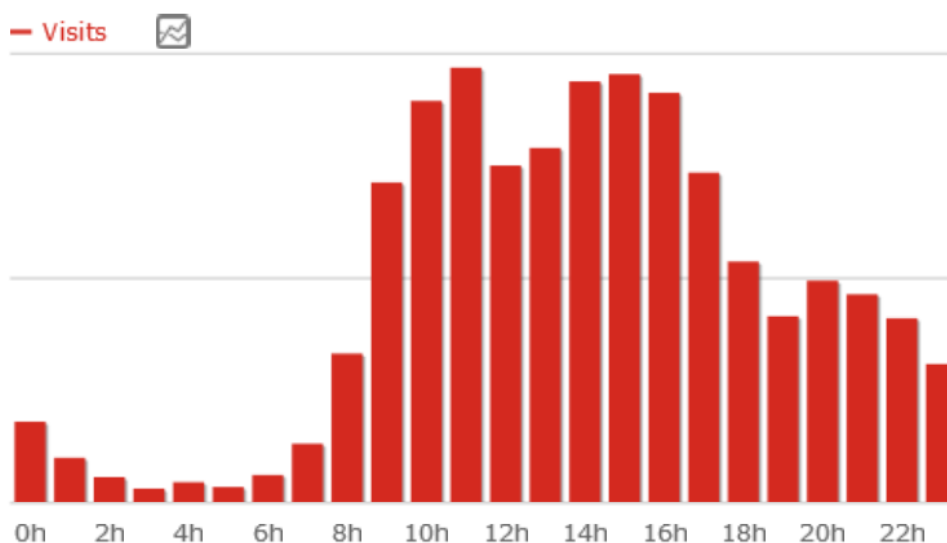
Actuellement, l'accès au site se fait essentiellement via des ordinateurs (88,9%). Parmi les autres modes de connexion, nous savons que 5,9% des visites sont réalisées à partir d'un smartphone et 4,6% via une tablette.

QUELLES SONT LEURS HABITUDES DE VISITE ?

Parmi toutes les visites effectuées sur le site, moins de la moitié (48,8%) durent plus de 10 secondes. Un peu plus d'un tiers des visites (32,4%) durent plus d'une minute, 9,2% durent plus de 10 minutes et enfin 2,3% durent plus de 30 minutes.

Nous pouvons également nous intéresser à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'un nombre de pages élevé peut être tant un signe de succès (site intéressant qui a su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information recherchée ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu). Pour le site www.cegesoma.be, environ 40% des visites comptabilisent trois pages vues ou plus, et 7,1% des visites comptent plus de 10 pages vues.

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 18h.



www.cegesoma.be - Fréquentation par heure (selon l'heure du serveur)

De plus nous pouvons également observer l'évolution des visites à travers le temps. De manière générale, nous observons une baisse de fréquentation régulière durant le week-end.



www.cegesoma.be – Évolution des visites, jour par jour

De plus, nous voyons un grand pic d'affluence autour du 2 février. En observant quelles sont les pages d'entrées (sur le site) les plus populaires à cette date-là, outre la homepage, nous pouvons identifier deux éléments qui sont probablement à l'origine de cette soudaine affluence : la publication d'une information concernant la visite de Elke Sleurs (secrétaire d'État à la Politique scientifique) au CegeSoma, ainsi qu'une annonce concernant l'annulation d'une conférence prévue le 19 février. Au final, il est apparu après concertation avec la responsable de la communication qu'il s'agit d'articles mis en avant dans la newsletter envoyée par l'institution. En effet, même si cette dernière n'est pas reconnue comme telle par Piwik, nous pouvons néanmoins observer des pics de fréquentation lors des envois de cette dernière.

D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour le Cegesoma, nous savons que 36% des visites se font en accès direct, c'est-à-dire que la personne a soit enregistré le site dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Plus de 50% des visites (53,3%) sont le fruit d'une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduit au site web du Cegesoma pourrait être très instructif. Malheureusement la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Nous avons ainsi plus de 12 000 mots-clés « not defined ». Les mots-clés disponibles ayant occasionné plus de 10 visites sont : « cegesoma » (138 visites) ; « soma » (56 visites) ; « ceges » (46 visites) ; « cegesoma.be » (22) et « <http://www.cegesoma> » (16).

Près de 10% des visites (9,6%) correspondent à un clic sur un lien vers le site du Cegesoma effectué sur d'autres sites web/réseaux sociaux. Nous retrouvons Facebook en tête de liste, cependant le détail est un peu confus et il n'est pas certain que tous les postes Facebook aient été pris en compte.

Enfin, la newsletter de l'institution est à l'origine de 1% des visites. Comme évoqué précédemment, cela s'explique par le fait que cette dernière n'a pas été configurée de manière à être reconnue par Piwik.

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE / SORTIE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ces forces et ces faiblesses.

En ce qui concerne les pages d'entrées les plus fréquentes, nous trouvons surtout la page d'accueil en français (21,8%), néerlandais (21,1%) et anglais (10,6%). Viennent ensuite les pages présentant les archives, en français (3,1%) et en néerlandais (1%).

En ce qui concerne les pages ayant un taux de sortie élevé, nous trouvons parmi celles-ci beaucoup de pages de présentation des membres du personnel de l'institution ou de pages liées à des événements. Or, il s'agit souvent de pages "cul-de-sac", ne renvoyant

généralement pas vers d'autres contenus. Il pourrait donc être intéressant de réfléchir à ce que l'on veut proposer au visiteur une fois qu'il est sur ce type de pages.

GOALS

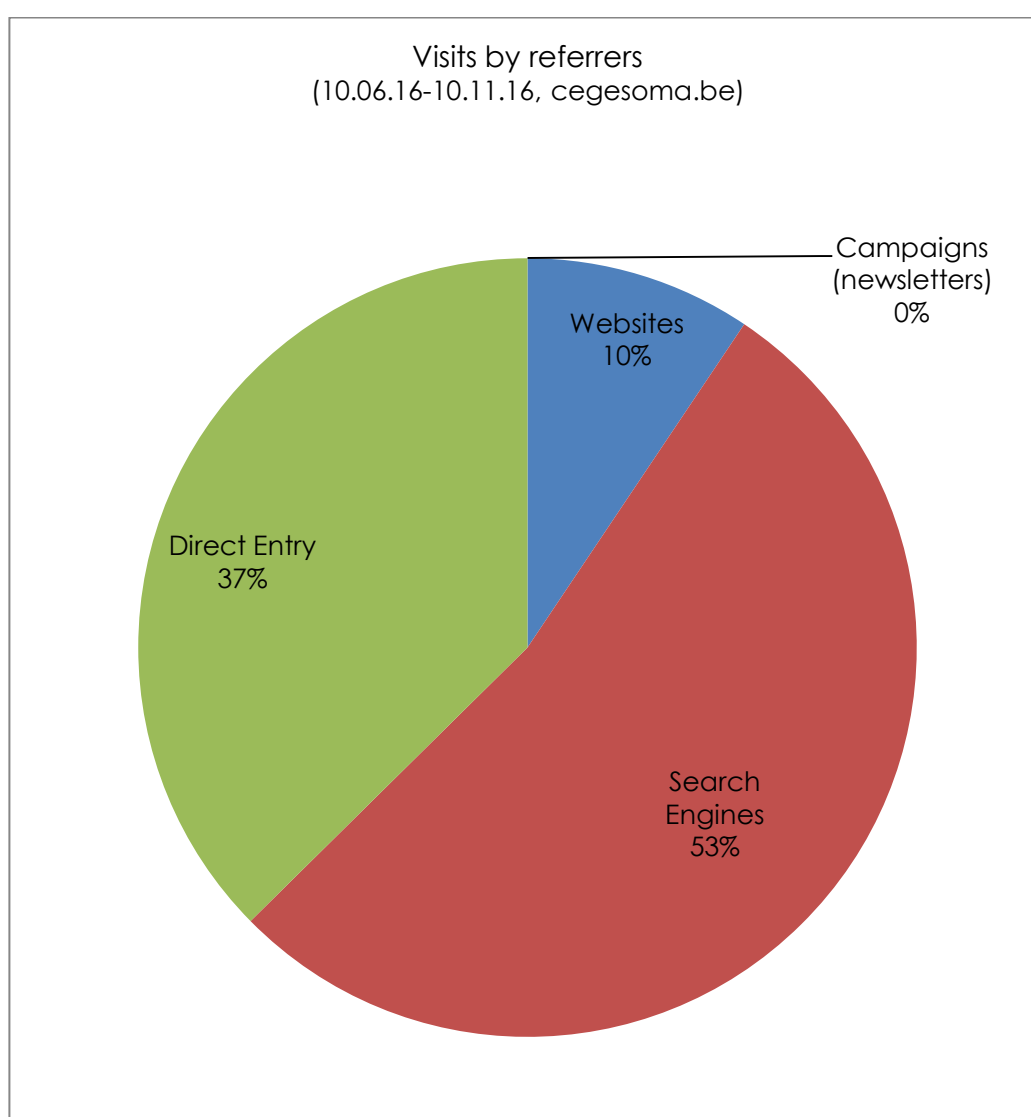
Pour le site cegesoma.be, 11 goals ont été identifiés :

| CEGESOMA.BE | | | |
|-------------|--|--|----------------------------------|
| Name | Access to | Web pages that contain | Taux de conversion ⁴⁵ |
| PRA | Practical information | Contact info, address, opening hours, prices of reading room cards | 6,8% |
| CONT II | Meta-information | Historical background, institution's mission and structure | 13,5% |
| CONT III | News and agenda | Agenda, news, newsletter subscription | 22,7% |
| TUT | Help and tutorial | FAQ, tutorials, site and search engine manuals, help pages, search assistance | 3,9% |
| VAL | Valorisation documents | Articles about collections | 6,9% |
| VAL-I | Valorisation websites | Hyperlinks to websites containing digital documents and virtual exhibitions created by the institution | 3,2% |
| CAT | Main catalogue | Hyperlinks to main digital catalogue | 18,8% |
| MET-PDF | Metadata related to the collections | Metadata related to the collections (outside main digital catalogue) | 2,0% |
| REP/DEP | Ordering reproductions / Deposit information | Information related to ordering of reproductions or the deposit of documents | 4,8% |
| REP-PDF | Ordering reproductions | Information related to ordering of reproductions (PDF) | 1,4% |
| SHARE | Sharing options | Possibility to send a link to a friend via this webpage | 0,2% |

⁴⁵ Moyenne.

Les taux de conversion présentés ici proviennent tous d'une période d'analyse de 5 mois (10.06.2016-10.11.2016)⁴⁶.

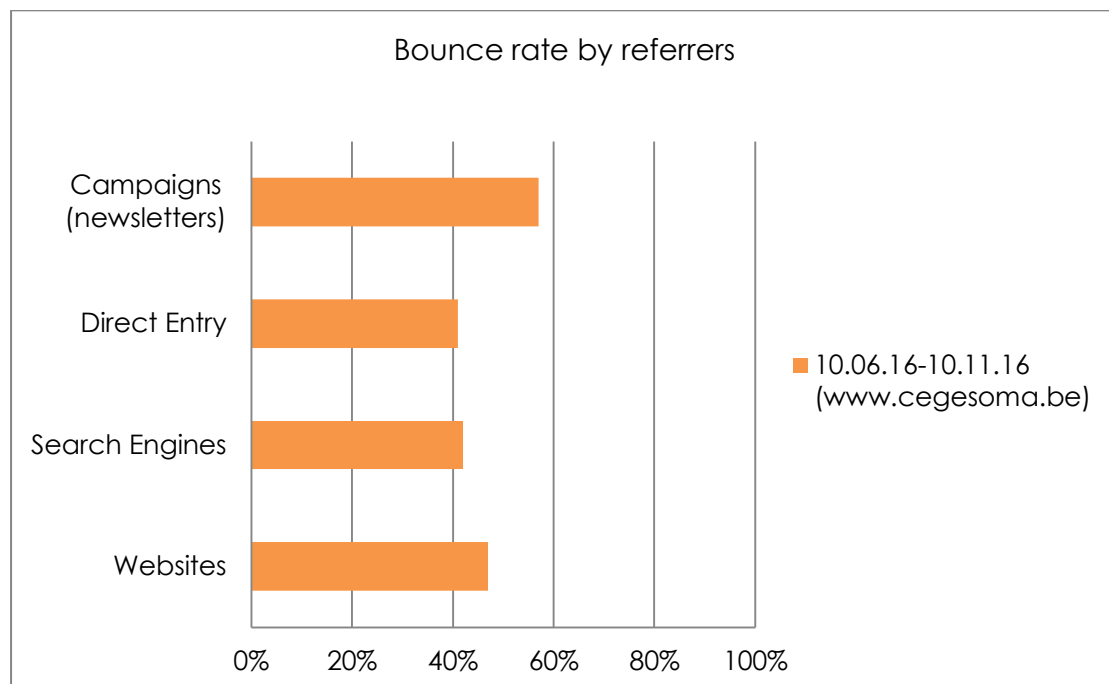
Durant ce laps de temps, 17 237 visites ont été comptabilisées. Avant de présenter les taux de conversion par goal, il semble important de d'abord se pencher sur l'origine de ces visites. Segmentation des visites effectuées entre le 10.06.2016 et le 10.11.2016 :



⁴⁶ Les goals, qui ne sont pas rétroactifs, ont été installés durant l'été 2016, au terme d'un long processus préparatoire. C'est pour cette raison que la période d'analyse n'est pas la même que pour la première partie du rapport.

Cela nous permet de relever les 0% correspondant aux visites générées par des clics sur un lien présent dans une newsletter. En réalité, 0% est la valeur arrondie pour 7 visites sur un total de 17 237. Cependant, sachant que cette somme reste minime et qu'il s'agit de surcroît de clics générés par la newsletter des Archives et non par celle des Archives, nous ferons abstraction de ce segment dans la suite de l'analyse.

Par ailleurs, si l'on veut se pencher sur les taux de conversion d'un site, il est utile de garder en tête le bounce rate de chaque segment (si l'on sait par exemple que plus de la moitié des utilisateurs quittent le site dès la page d'accueil, alors un taux de conversion de 20% semble directement plus encourageant).



PRA = Information pratique = 6,8%

Cela signifie que 6,8% des visites (c'est-à-dire 1 774) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

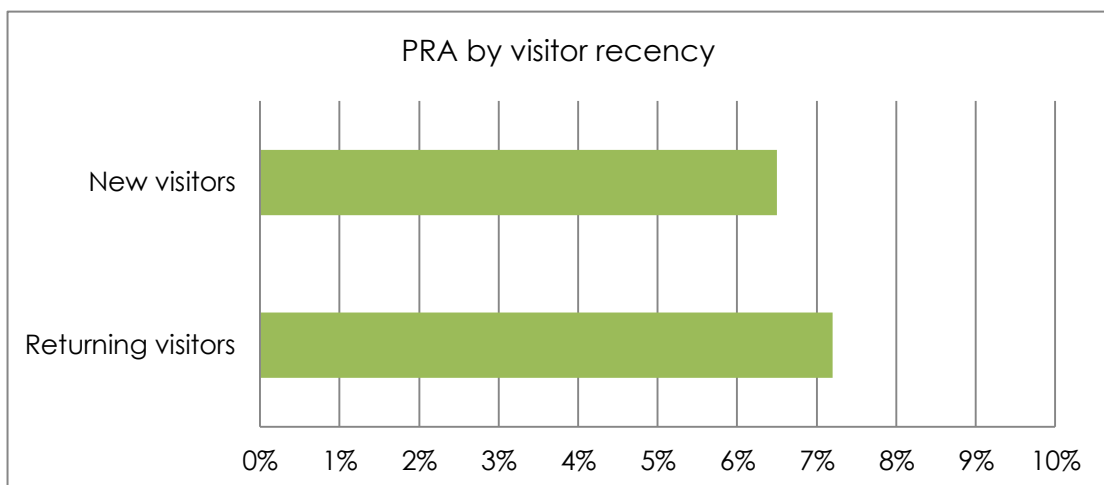
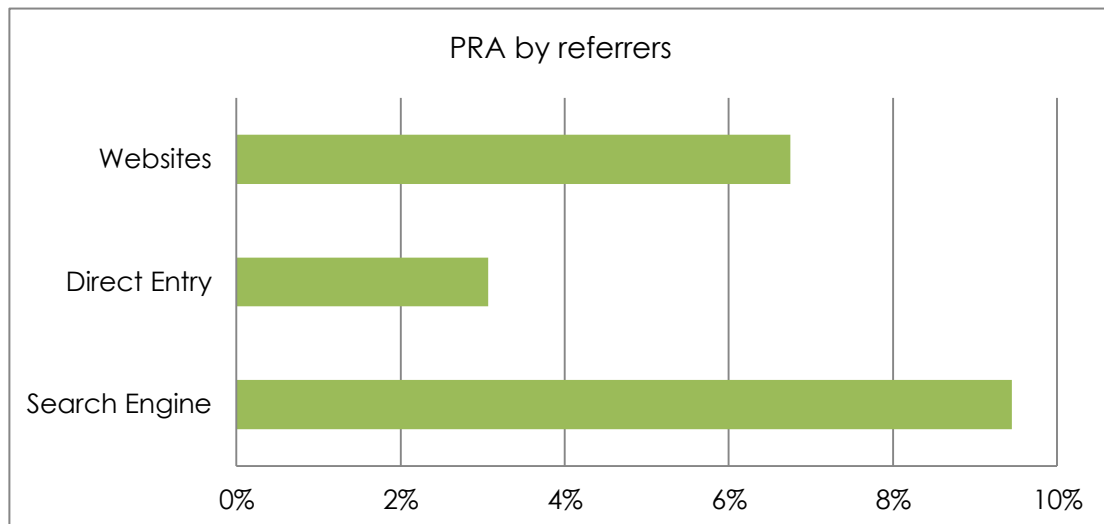
http://www.cegesoma.be/cms/sallelecture_fr.php

http://www.cegesoma.be/cms/contacts_fr.php

http://www.cegesoma.be/cms/localisation_fr.php

http://www.cegesoma.be/cms/plaintes_fr.php

Taux de conversion par segment⁴⁷



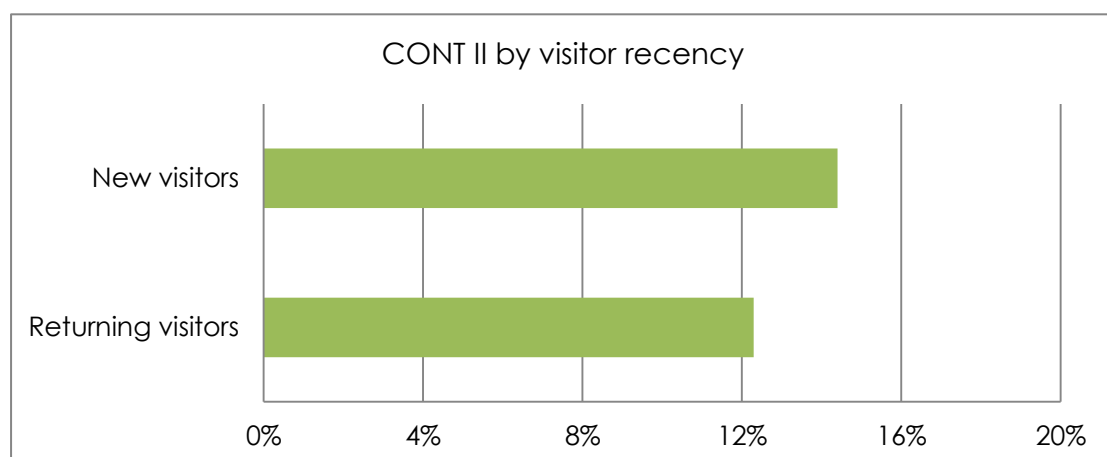
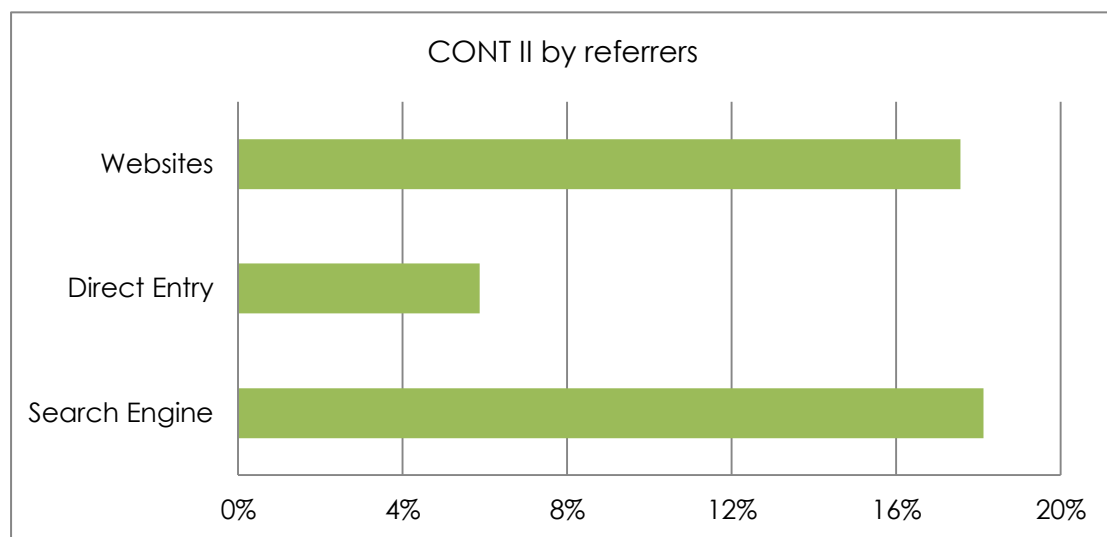
⁴⁷ Le recours à d'autres segments est bien sûr envisageable !

CONT II = Meta-information = 13,5%

Cela signifie que 13,5% des visites (2 327) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/cms/historique_fr.php
http://www.cegesoma.be/cms/rech_encours_fr.php
http://www.cegesoma.be/cms/projpubli_fr.php
http://www.cegesoma.be/cms/proj_clotures_fr.php
http://www.cegesoma.be/cms/bibliotheque_fr.php
http://www.cegesoma.be/cms/depodoc_fr.php
http://www.cegesoma.be/cms/amisceges_fr.php
http://www.cegesoma.be/cms/benevol_fr.php
http://www.cegesoma.be/cms/mission_fr.php

Taux de conversion par segment



CONT III = Nouvelles et agenda = 22,7%

Cela signifie que 22,7% des visites (3 915) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/cms/maillinglist_fr.php

http://www.cegesoma.be/cms/agenda_fr.php

http://www.cegesoma.be/cms/activ_scientif_fr.php

http://www.cegesoma.be/cms/activ_scientif_fr.php?article=3008

http://www.cegesoma.be/cms/activ_scientif_fr.php?article=3019&pagnbr=6&pagofs=0

http://www.cegesoma.be/cms/conferences_fr.php

http://www.cegesoma.be/cms/nouvelles_fr.php

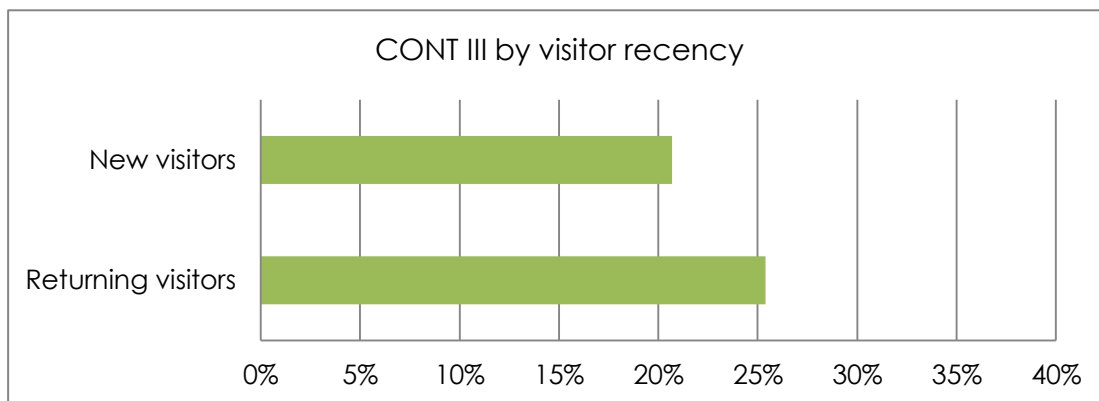
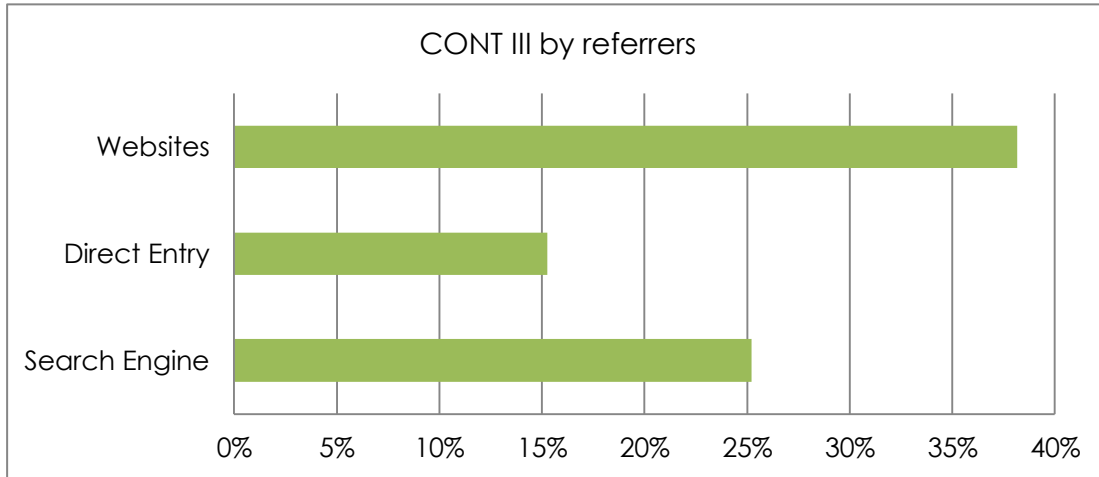
http://www.cegesoma.be/cms/expo_fr.php

http://www.cegesoma.be/cms/index_fr.php?article=2445

Etc.⁴⁸

⁴⁸ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires.

Taux de conversion par segment



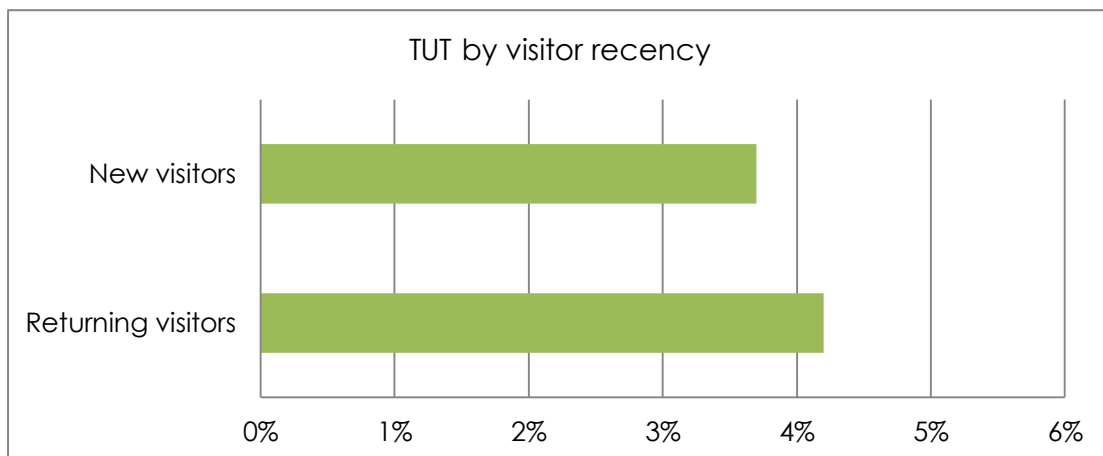
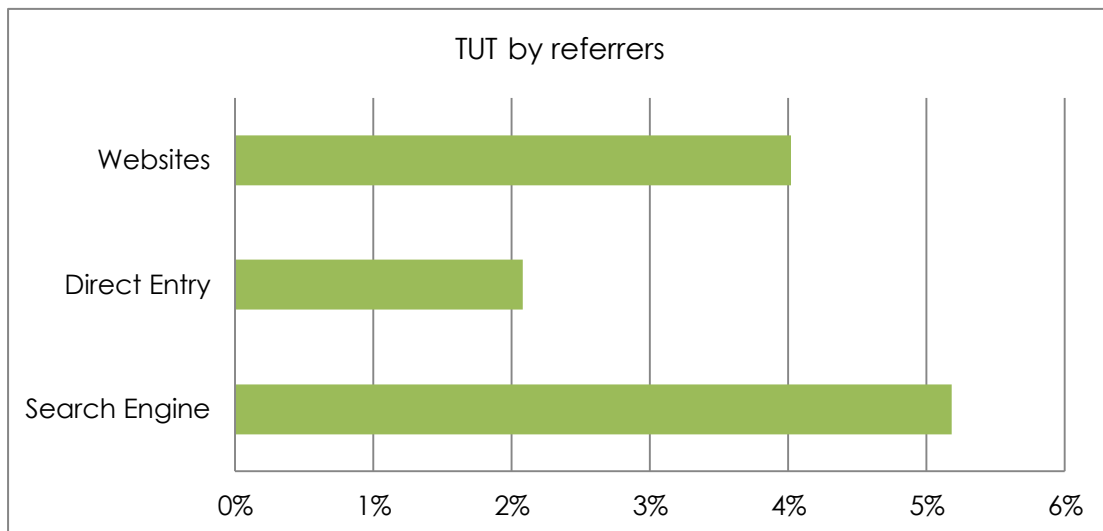
TUT = Aide et tutoriels = 3,9%

Cela signifie que 3,9% des visites (674) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/cms/banqueimages_fr.php?

http://www.cegesoma.be/cms/cataldivers_fr.php?cnvis=109365

Taux de conversion par segment

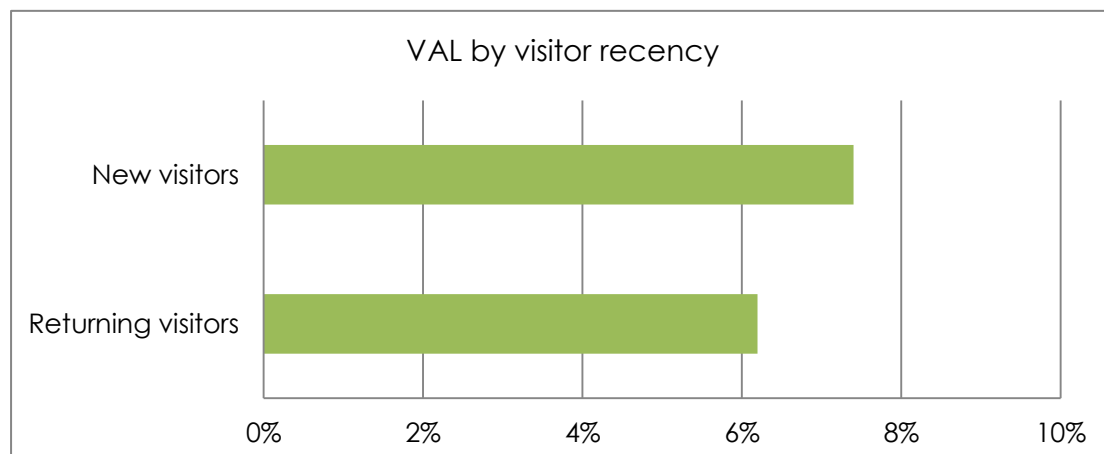
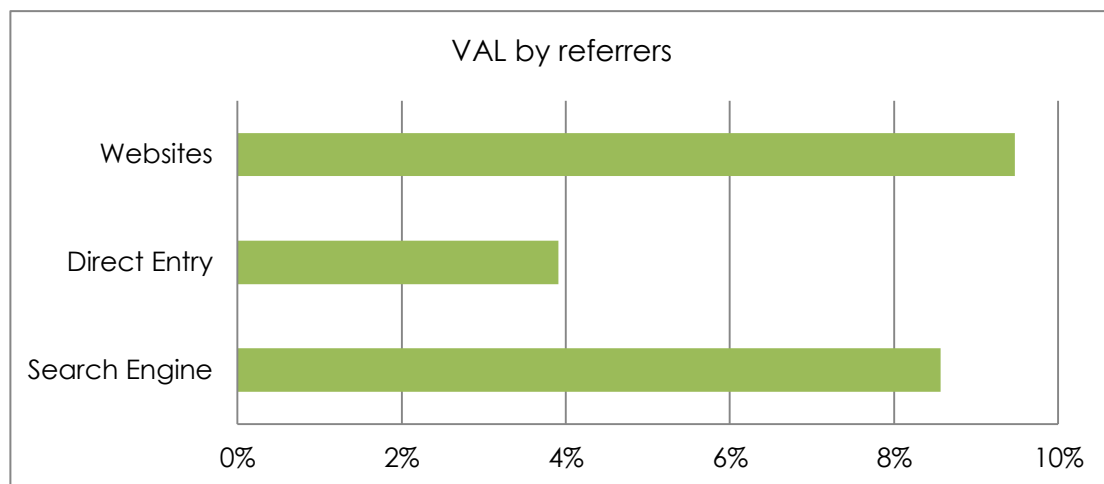


VAL = valorisation (documents) = 6,9%

Cela signifie que 6,9% des visites (1192) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

- http://www.cegesoma.be/cms/belgwarpress_fr.php
- http://www.cegesoma.be/cms/belgwarpress_fr.php?article=2593
- http://www.cegesoma.be/cms/belgwarpress_fr.php?article=3005
- http://www.cegesoma.be/cms/sitesweb_fr.php?article=2703
- http://www.cegesoma.be/cms/sitesweb_fr.php?article=2535
- http://www.cegesoma.be/cms/sitesweb_fr.php?article=2618
- http://www.cegesoma.be/cms/sitesweb_fr.php?article=2703
- http://www.cegesoma.be/cms/resistflandre_fr.php?cnvis=984511
- http://www.cegesoma.be/cms/autres_publi_fr.php
- http://www.cegesoma.be/cms/sitesweb_fr.php?article=3097

Taux de conversion par segment



VAL-I = valorisation (sites-web) = 3,2%

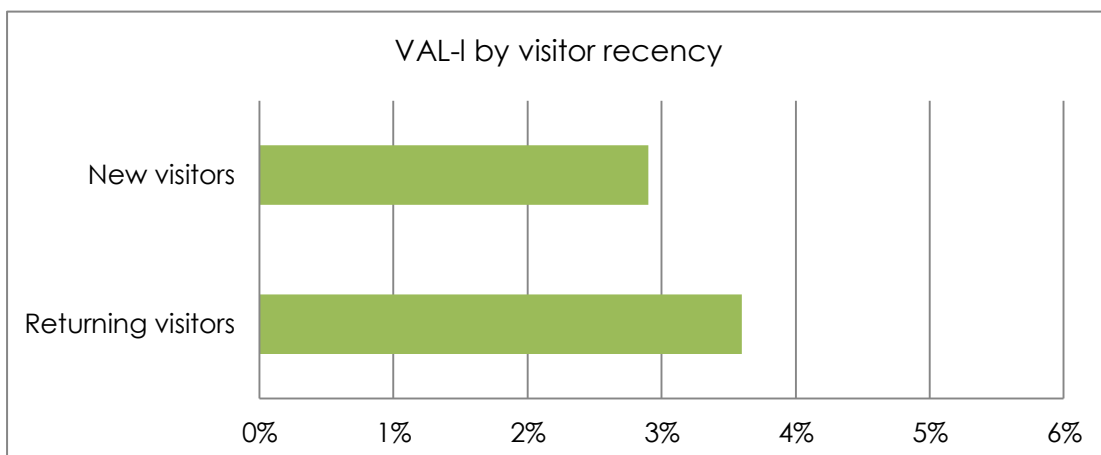
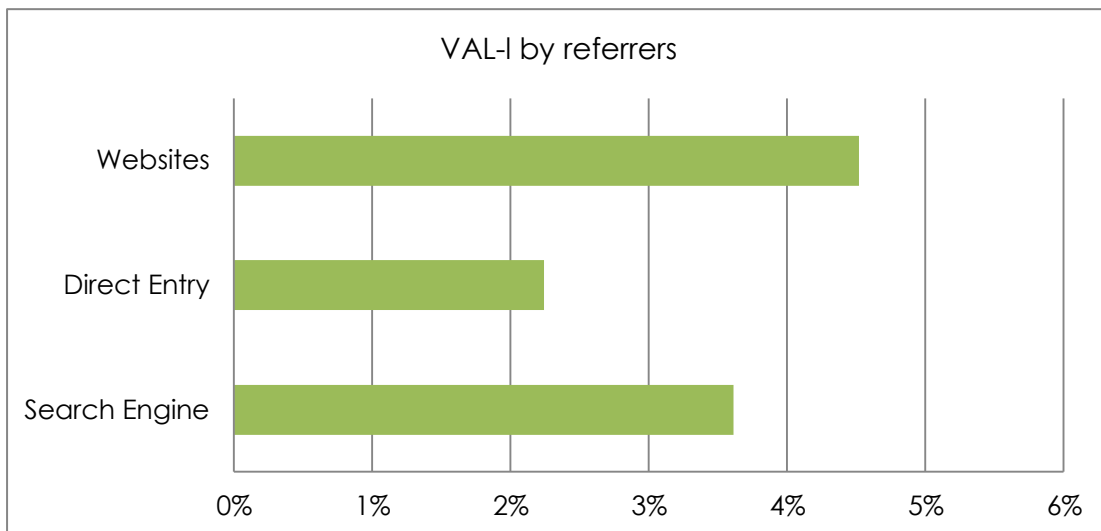
Cela signifie que 3,2% des visites (548) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

<http://www.journalbelgianhistory.be/fr>

<http://www.14-18.bruxelles.be/index.php/fr/>

<http://warpress.cegesoma.be>

Taux de conversion par segment



CAT = catalogue principal = 18,8%

Cela signifie que 18,8% des visites (3 246) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

<http://opac.cegesoma.be/fr/archview/list>

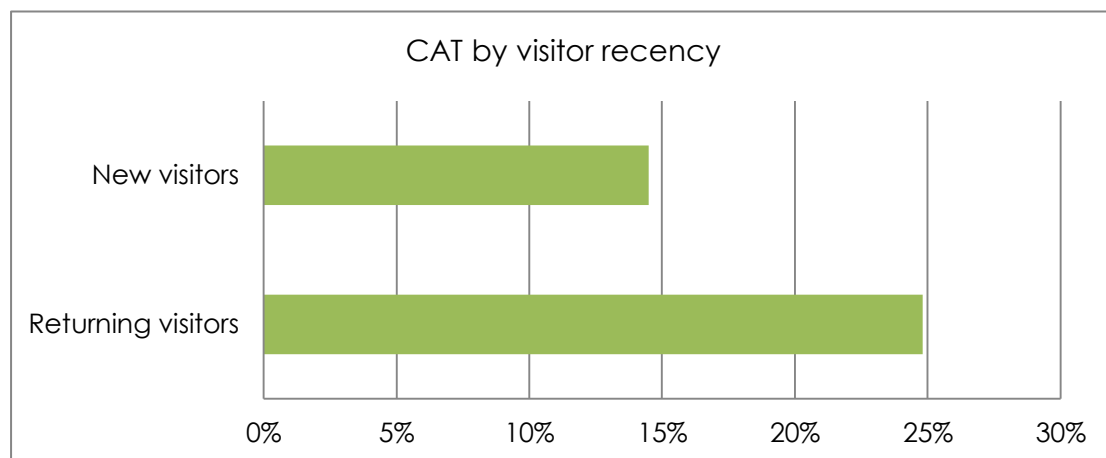
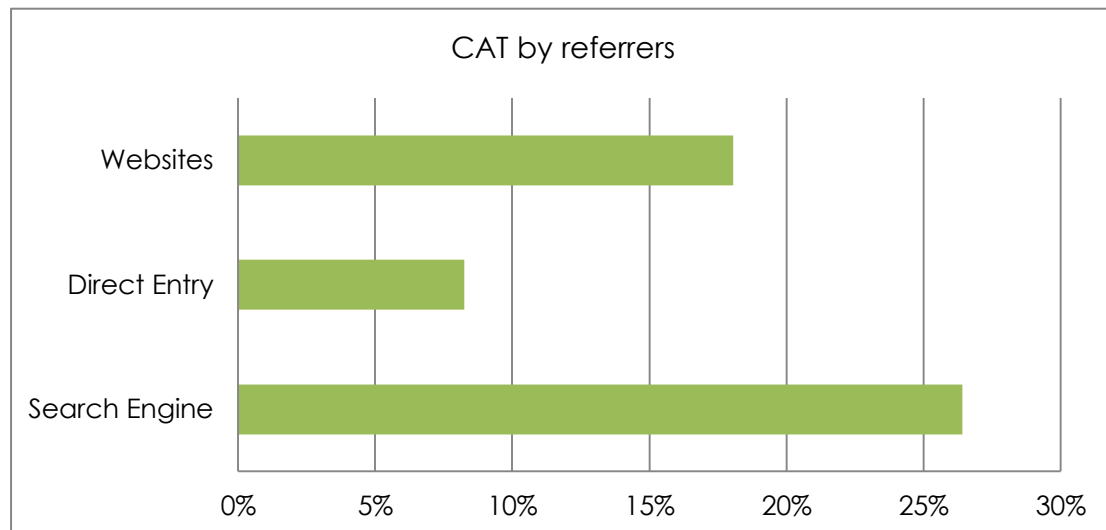
http://www.pallas.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac_viv.htm

<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&rn=790&htdoc=general/opac.htm>

<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&rn=791&htdoc=general/opac.htm>

Etc.⁴⁹

Taux de conversion par segment



⁴⁹ Grâce à l'utilisation d'expressions régulières, il suffit de connaître les motifs récurrents pour pouvoir prendre en compte toutes les URLs similaires composée de pallas.cegesoma.be... ou opac.cegesoma.be....

MET-PDF = 2,0%

Cela signifie que 2,0% des visites (341) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Affiches.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_DVD_series.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_DVD_titres.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_VieQuot_noms2.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_VieQuot_sujets2.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Interviews_noms.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Interviews_sujets.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Interviews_noms.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/JoursGuerre_radio_AA1593.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Jours_de_guerre_DVD_juin_2014_site.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/JoursGuerre_TV_AA_1450.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Affiches.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Radio_dates.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Radio_sujets.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/AA_1321_RadioBruxelles.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_DVD_series.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_DVD_titres.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/AA1825_VRT_Interviews_C_R_C_R_A_V.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Memoires.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Radio_dates.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Radio_sujets.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/AA_1321_RadioBruxelles.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/AA1825_VRT_Interviews_C_R_C_R_A_V.pdf
http://www.cegesoma.be/docs/media/TrouvDoc/catalogues/Catal_Memoires.pdf
http://www.cegesoma.be/docs/media/Bibliographies/biblio_1996.pdf

http://www.cegesoma.be/docs/media/Bibliographies/Bibliography_ESF.pdf

http://www.cegesoma.be/docs/media/Bibliographies/bibliografie_verzet.pdf

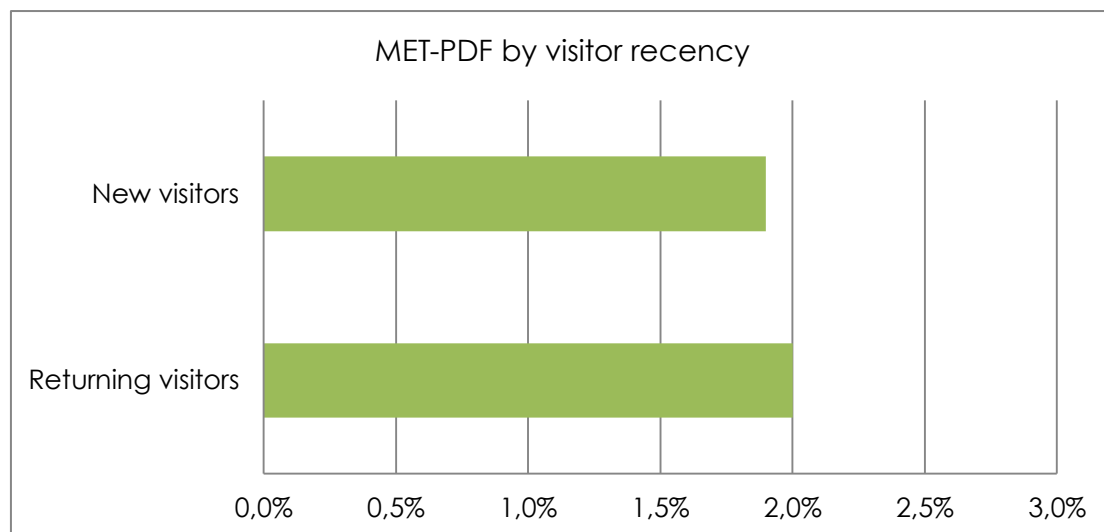
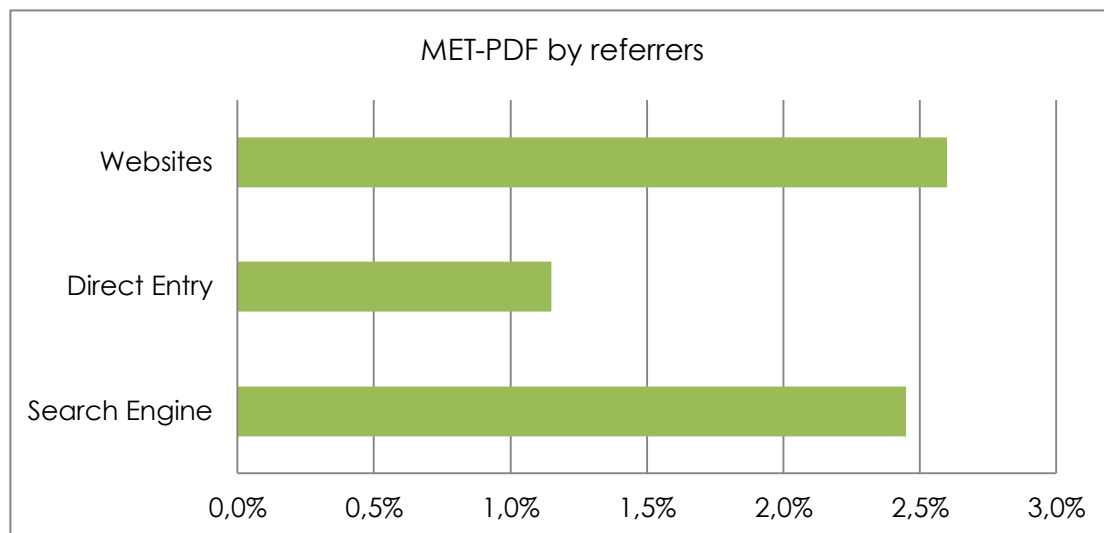
http://www.cegesoma.be/docs/media/Bibliographies/biblio_1970_1980.pdf

http://www.cegesoma.be/docs/images/stories/ceges/Collections/lijst_lopende_periodieken_2015.pdf

http://www.cegesoma.be/docs/images/stories/ceges/Collections/lijst_lopende_periodieken_2015.pdf

http://www.cegesoma.be/docs/Invent/AB_Lijst.pdf

Taux de conversion par segment



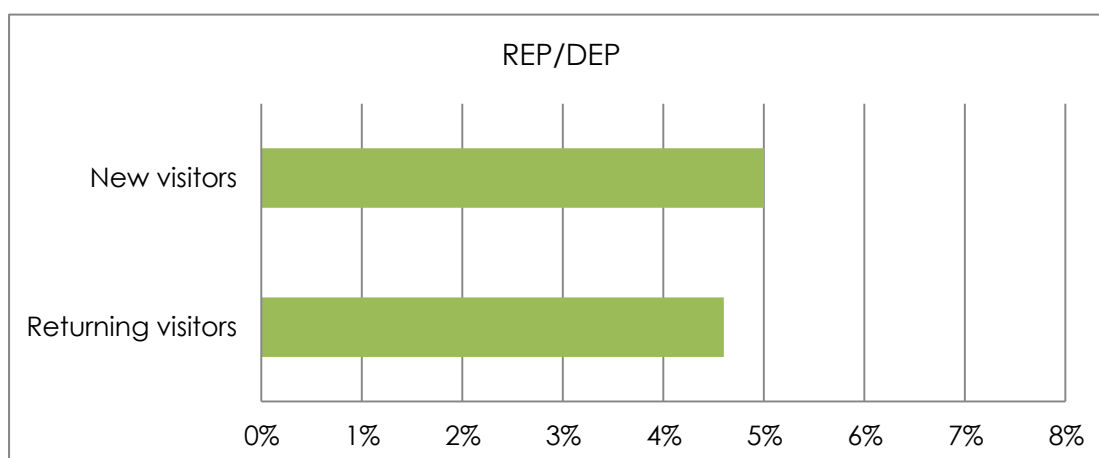
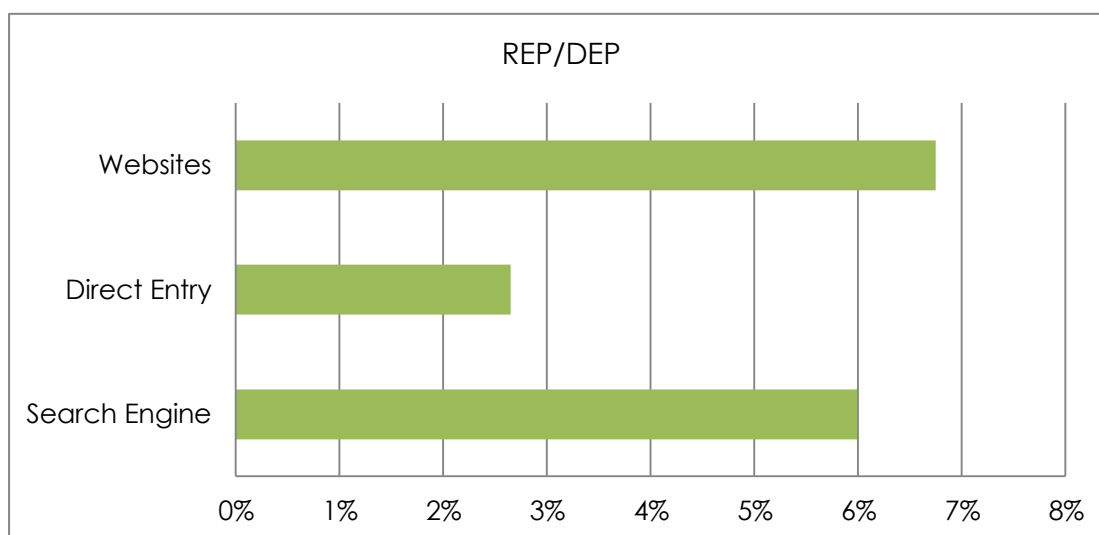
REP/DEP = reproduction de documents /dons-dépôts = 4,8%

Cela signifie que 4,8% des visites (829) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/cms/reproduction_fr.php

http://www.cegesoma.be/cms/depodoc_fr.php

Taux de conversion par segment



REP-PDF = reproductions de documents pdf = 1,4%

Cela signifie que 1,4% des visites (245) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/docs/media/Phototheque/TarifsPhotos_fr_Juin2014.pdf

http://www.cegesoma.be/docs/media/Phototheque/BonCommande_Fr_Juin2014.pdf

http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/Tarifs_mai2014_fr.pdf

http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/TarifsInterviews_juin2014.pdf

http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/Tarieven_mai2014_nl.pdf

http://www.cegesoma.be/docs/media/Phototheque/TarievenFotos_nl_juillet2014.pdf

http://www.cegesoma.be/docs/media/Phototheque/Bestelbon_Nl_juillet2014.pdf

http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/TarifsInterviews_juin2014_nl.pdf

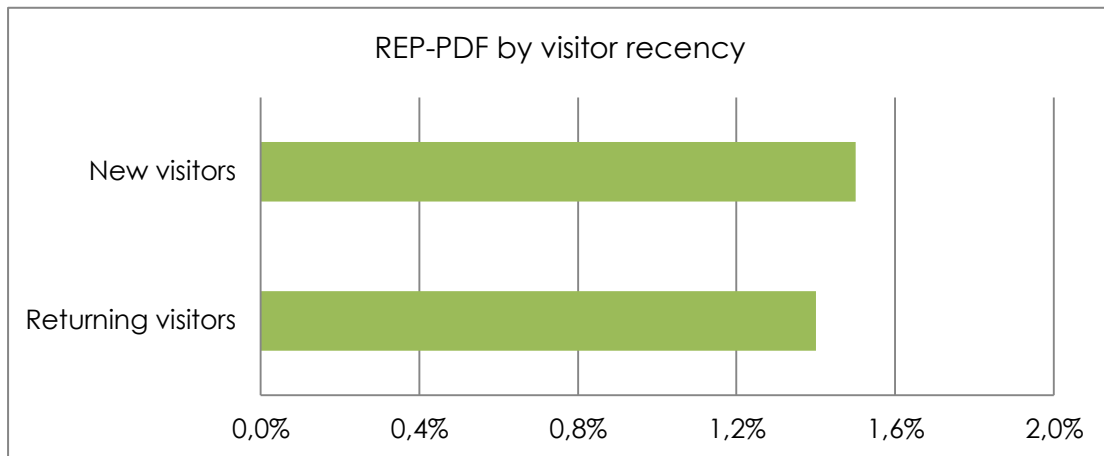
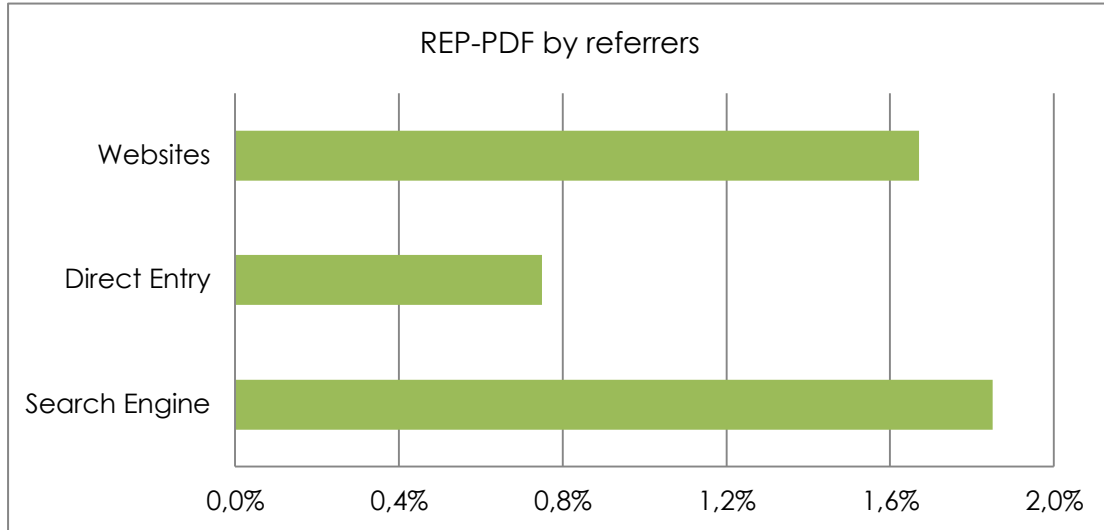
http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/Tarifs_mai2014_en.pdf

http://www.cegesoma.be/docs/media/Phototheque/PricePhotos_Juin2014.pdf

http://www.cegesoma.be/docs/media/Phototheque/OrderForm_En_June2014.pdf

http://www.cegesoma.be/docs/media/InfoPratiques/ReproDocu/TarifsInterviews_juin2014_en.pdf

Taux de conversion par segment



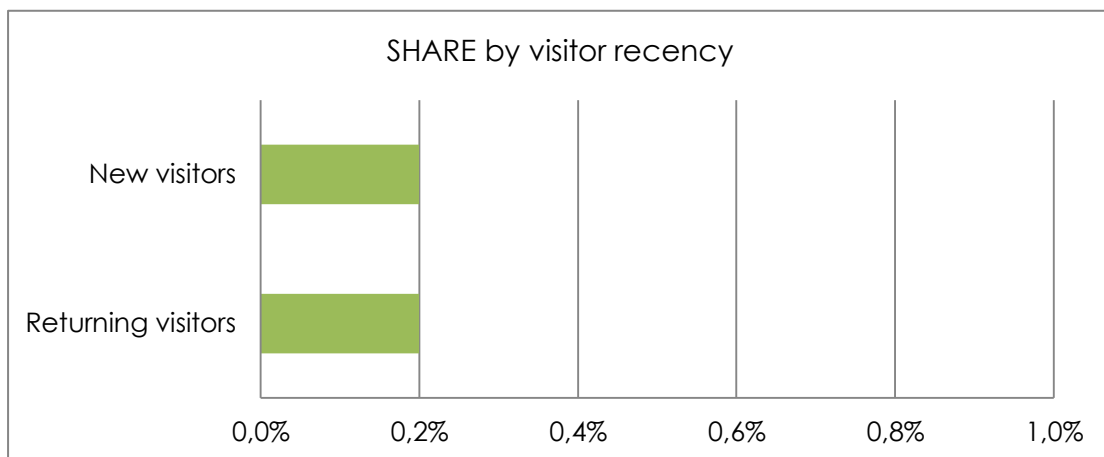
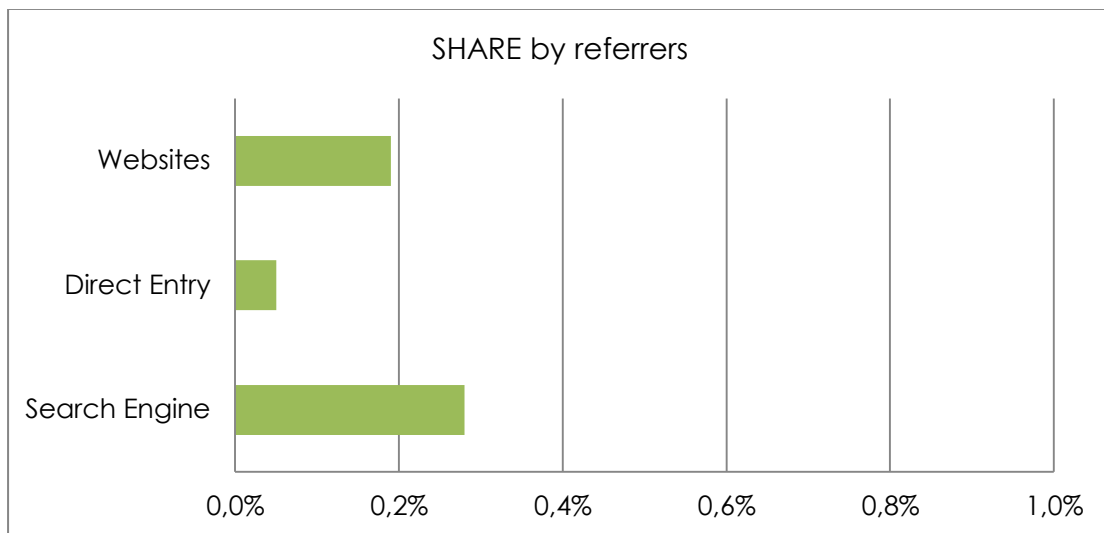
SHARE = options de partage = 0,2%

Cela signifie que 0,2% des visites (32) comportent au moins un passage sur l'une de ces pages (ou sur l'une des pages équivalentes en néerlandais ou anglais) :

http://www.cegesoma.be/cms/envoiami_en.php?

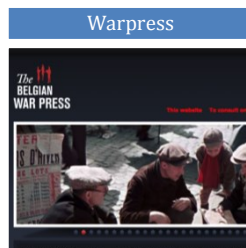
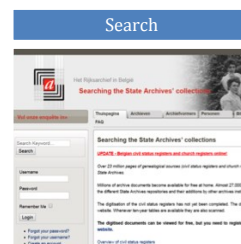
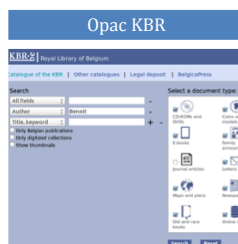
Etc.⁵⁰

Taux de conversion par segment



⁵⁰ L'expression régulière permet de prendre en compte tous les tyoes de pages comportant la mention d'envoi à un ami.

Partie II : sites des catalogues en ligne, sites d'accès aux collections numérisées et sites de valorisation des collections



Introduction

Cette seconde partie se concentre autour des catalogues en ligne, des sites d'accès aux collections numérisées et des sites de valorisation. Nous présentons d'abord la méthodologie propre aux catalogues (Introduction), vient ensuite une vue comparative (2.1 pour les catalogues en ligne et sites d'accès aux collections numérisées, 2.2 pour les sites de valorisation des collections) et enfin des analyses spécifiques (2.3).

En ce qui concerne la méthodologie, il faut savoir que la structure et les caractéristiques des catalogues en ligne sont telles qu'elles que l'interface graphique Piwik ne peut nous fournir qu'un nombre limité d'informations. Si l'on souhaite mieux cerner l'usage qui est fait d'un catalogue, une approche manuelle s'avère donc indispensable. Par manuelle, il faut comprendre que les différentes étapes de collecte et traitement des données sont réalisées à l'aide de scripts Python⁵¹ rédigés à la main, plutôt que par le biais de l'outil clé en main de Piwik.

Cette méthode a pour inconvénient de nécessiter davantage de temps de préparation en amont et d'être moins facilement transférable au personnel travaillant dans les institutions partenaires du projet. Cependant, elle permet d'aller bien au-delà des limites que présente l'interface Piwik et d'obtenir des données riches en informations, telles que les mots-clés encodés par les utilisateurs dans un moteur de recherche. Par ailleurs, un grand nombre d'opérations peut être automatisé, rendant possible l'analyse de corpus très imposants, comme c'est par exemple le cas avec le catalogue Search.

Étant donné l'important volume de données à traiter, la diversité des sites web faisant partie du projet et le relativement court laps de temps à disposition, l'accent a été mis sur quatre questions de recherche :

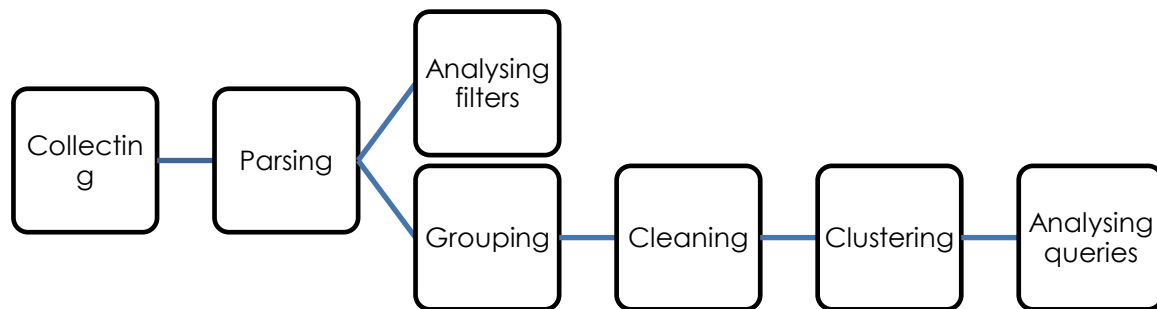
- **quel est le profil des visiteurs ?**
- **quel est le ratio de « recherche simple » versus « recherche avancée » ?**
- **que peut-on savoir concernant l'usage des filtres ?**

⁵¹ Il s'agit d'une suite de commandes rédigés à l'aide du langage de programmation « Python ». Ces commandes permettent le traitement automatisé de grandes quantités de données.

- quelles sont les thématiques les plus récurrentes parmi les requêtes des utilisateurs ?

Bien entendu, l'exploration pourrait se poursuivre en vue de répondre à d'autres questions. De plus, il pourrait être intéressant de croiser les résultats obtenus afin d'établir une typologie des utilisateurs basée sur les caractéristiques de leurs parcours de visite.

En ce qui concerne la méthodologie utilisée, celle-ci peut être décomposée en plusieurs étapes. Un bref résumé est présenté ici. Pour tout ce qui relève de l'analyse des requêtes des utilisateurs, une description plus détaillée est proposée dans une publication annexe⁵².



La première étape a consisté à collecter et extraire les données en interrogeant à l'aide du langage MySQL les bases de données regroupant l'ensemble des informations sur l'activité des différents sites Web. Les URLs ont ensuite été utilisées pour puiser l'information permettant de répondre aux différentes questions de recherche. Afin de limiter la taille des fichiers à exploiter (certains font plusieurs Go), nous avons dû faire un tri pour ne garder que les URLs les plus pertinentes. Un second tri a ensuite été réalisé au sein des URLs elles-mêmes pour ne garder que certaines parties (on parle de « parsing ») stockées ensuite de façon structurée afin de pouvoir automatiser et accélérer l'analyse. Ces étapes ont été réalisées à l'aide de scripts Python et de la librairie Pandas⁵³.

⁵² Chardonens, Anne, Hengchen Simon, Text Mining for User Query Analysis. A 5-Step Method for Cultural Heritage Institutions, Proceedings of ISI 2017 – the 15th International Symposium of Information Science, Humboldt-Universität zu Berlin, Allemagne, 13-15 mars 2017 (à paraître).

⁵³ Expliquer

Les fichiers résultants de ces différentes opérations ont pu servir ensuite, grâce à une analyse dans Open Refine, à comprendre de quelle manière et dans quelle mesure les utilisateurs se servent des paramètres de recherche avancée dans les catalogues des institutions.

Afin de pouvoir explorer le contenu des requêtes elles-mêmes, nous avons d'abord complété le script Python afin de ne garder qu'une occurrence par requête distincte, par visite. Des étapes de normalisation et de clustering ont ensuite été nécessaires afin de pouvoir associer les requêtes similaires. Enfin, l'analyse s'est terminée en ouvrant le fichier sortant (en format CSV) au sein d'OpenRefine.

Précisons encore que tous les résultats rassemblés dans cette section proviennent d'un échantillon d'analyse. Sauf indication contraire, ce dernier est constitué de toutes les données (de fréquentation des sites) qui ont été collectées entre le **01.01.2016** et le **01.07.2016**.

2.1. Vue comparative : sites des catalogues en ligne et sites d'accès aux collections numérisées

2.1.1. Visites et visiteurs

Chaque fois qu'un visiteur vient sur un site pour la première fois, ou qu'il visite une page plus de 30 minutes⁵⁴ après sa dernière page vue, une nouvelle visite est enregistrée⁵⁵. Un « visiteur » unique peut donc réaliser plusieurs visites par jour. À sa première visite, le visiteur est considéré comme un « New Visitor ». Dès sa seconde visite, il sera considéré comme un « Returning visitor ». Un même visiteur est reconnu grâce à l'usage de « cookies ».

⁵⁴ Cette durée, fixée par défaut à 30 minutes, peut être modifiée au sein de Piwik.

⁵⁵ Il faut noter que parmi toutes les visites enregistrées, il est possible que certaines aient été effectuées par des robots.

| | Search ⁵⁶ | OpacKBR ⁵⁷ | Pallas ⁵⁸ | Belgica ⁵⁹ | Warpress ⁶⁰ |
|--|----------------------|-----------------------|----------------------|-----------------------|------------------------|
| Visites JANVIER | 196 797 | 13 122 | 1 638 | 5 164 | 1 618 |
| Visiteurs uniques JANVIER | 48 570 | 7 247 | 1 055 | 3 989 | 1 255 |
| Visites FÉVRIER | 197 185 | 14 510 | 1 586 | 5 306 | 1 896 |
| Visiteurs uniques FÉVRIER | 48 385 | 7 943 | 1 111 | 3 996 | 1 464 |
| Visites MARS | 187 825 | 13 830 | 1 589 | 5 093 | 2 291 |
| Visiteurs uniques MARS | 46 584 | 7 708 | 1 039 | 3 978 | 1 776 |
| Visites AVRIL | 169 592 | 13 238 | 1 461 | 4 709 | 1 796 |
| Visiteurs uniques AVRIL | 44 113 | 7 173 | 976 | 3 532 | 1 248 |
| Visites MAI | 157 946 | 12 173 | 1 486 | 4 457 | 1 473 |
| Visiteurs uniques MAI | 41 953 | 6 565 | 944 | 3 387 | 1 131 |
| Visites JUIN | 149 715 | 11 215 | 1 324 | 3 713 | 976 |
| Visiteurs uniques JUIN | 39 165 | 5 888 | 862 | 2 915 | 788 |
| Total Visites (janvier à juin 2016) | 803 350 | 78 510 | 9 151 | 28 549 | 10 075 |
| Total pages vues (janvier à juin 2016) | 57 205 513 | 2 200 924 | 182 067 | 66 332 | 84 227 |

2.1.2. Répartition des visites

Les visites peuvent également être analysées de façon plus détaillée en les visualisant par segments. Par exemple, il est possible de connaître l'origine du trafic sur le site (on parle de « referrers ») et de répartir les visites selon cette variable-là. Il faut cependant noter que ce classement est réalisé automatiquement par Piwik. Ainsi, le segment « campaigns », qui peut par exemple prendre en compte les visites générées par un clic sur une newsletter, ne comporte aucune visite pour certains catalogues. C'est le cas par exemple des newsletters de la KBR qui ne sont pas reconnues comme telles par Piwik.

⁵⁶ <http://search.arch.be/>

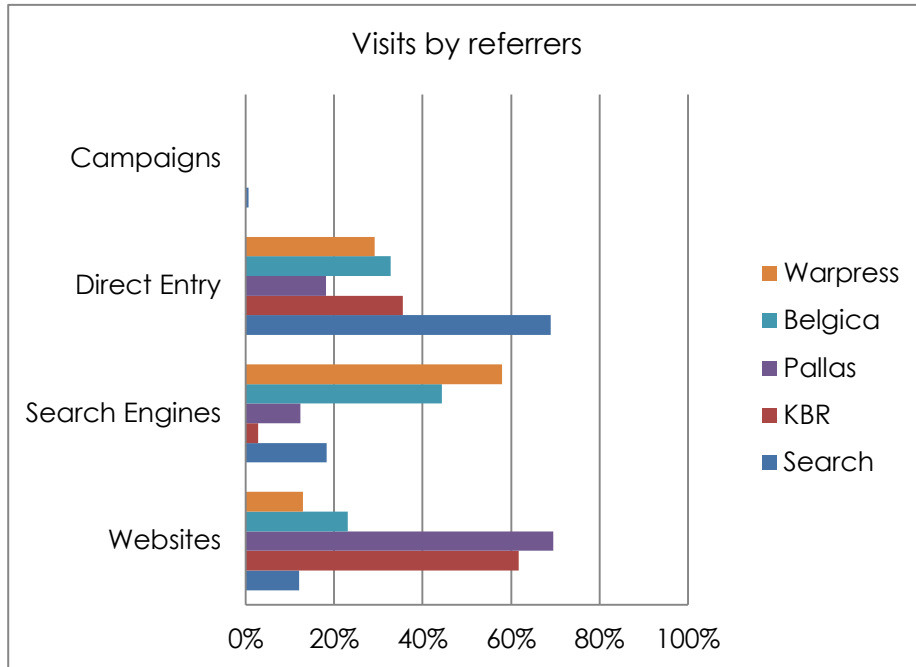
⁵⁷ <http://opac.kbr.be/>

⁵⁸ <http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htm>

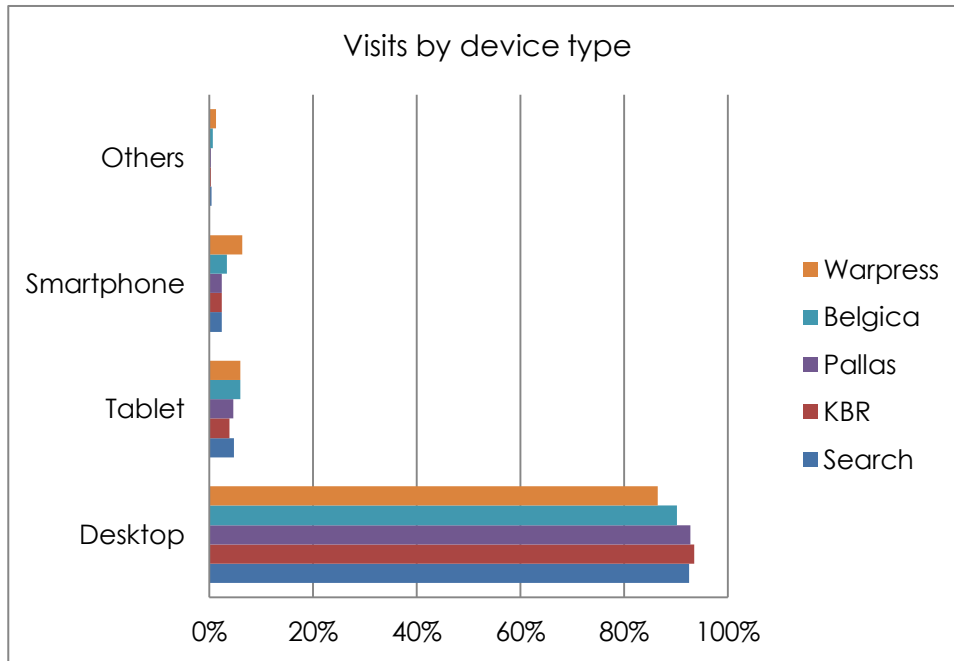
⁵⁹ <http://belgica.kbr.be/>

⁶⁰ <http://warpress.cegesoma.be/>

Répartition des visites par type d'accès

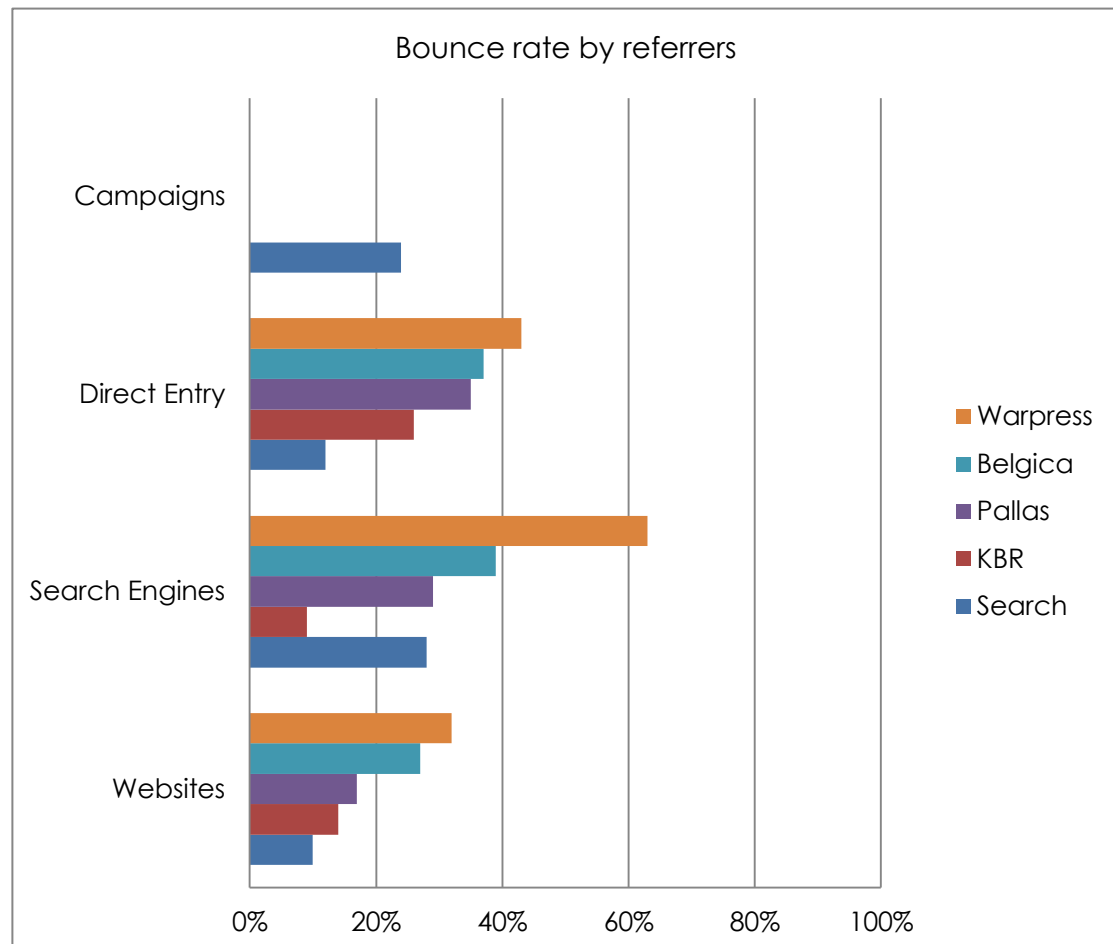
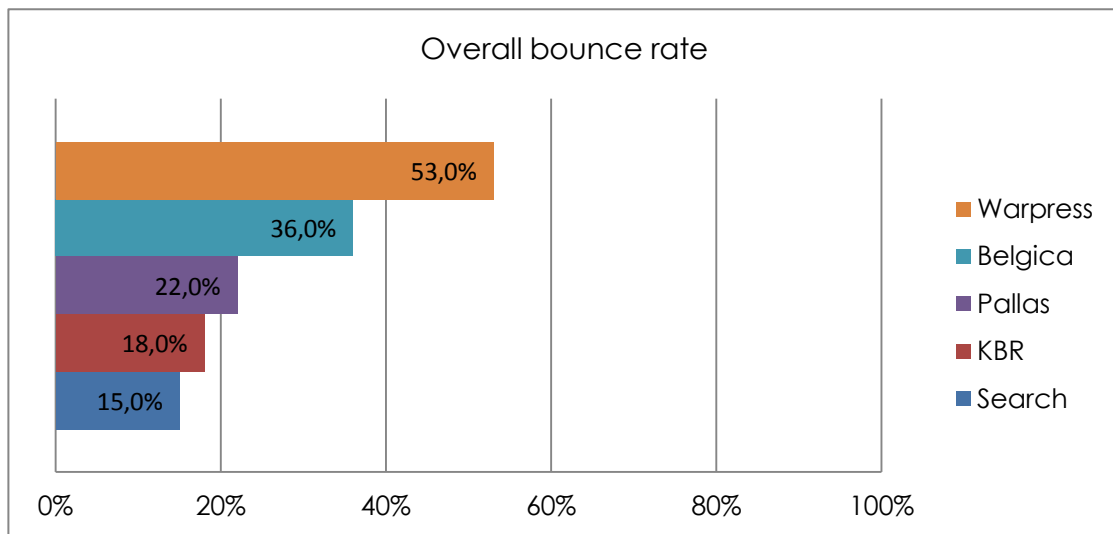


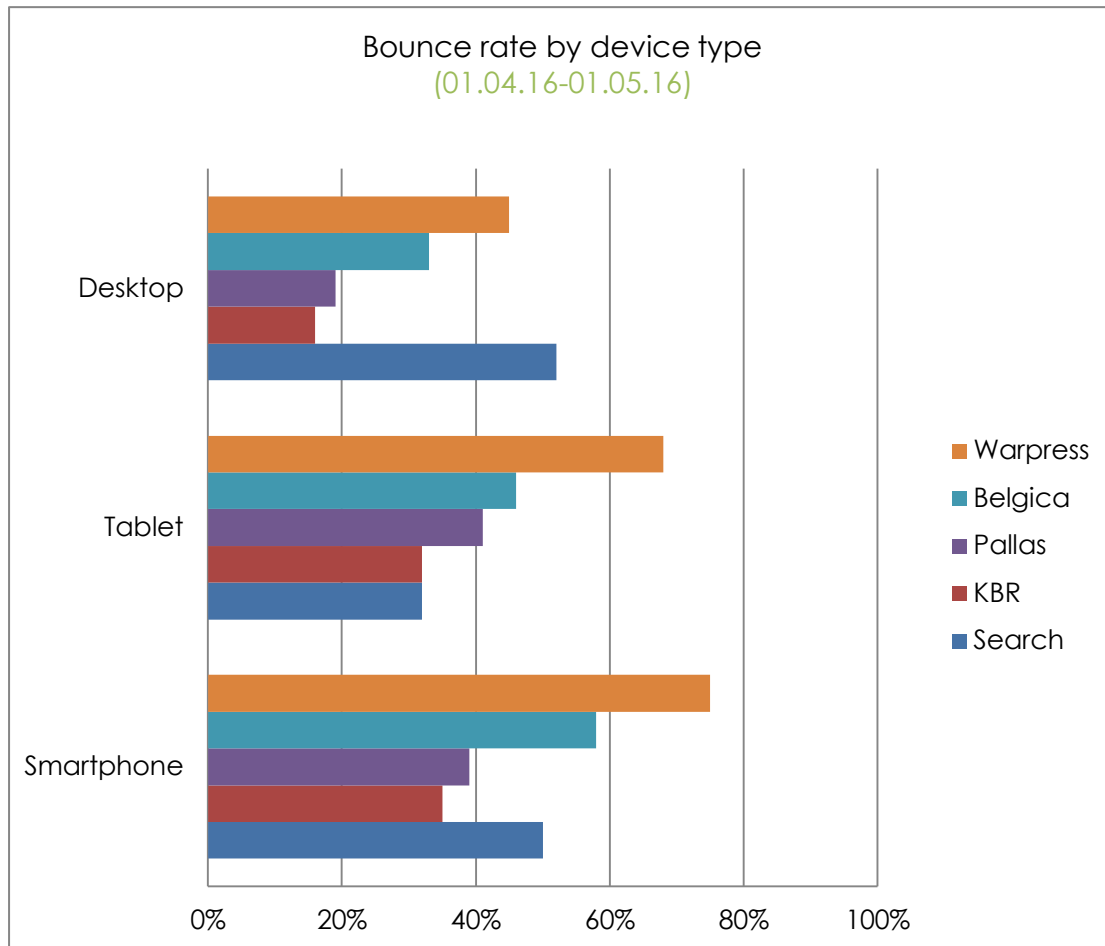
Répartition des visites par type de supports



Bounce rate

Il s'agit du taux de visites qui ne comptabilisent qu'une seule page vue car le visiteur a directement quitté le site (cela peut être la page d'accueil du site comme une autre page). Si l'on ignore les raisons précises de ce départ précipité, on peut en revanche détailler ce bounce rate par segments pour essayer d'en savoir plus.





Pour des raisons pratiques, la période d'analyse a été exceptionnellement réduite à un mois (01.04.16-01.05.16) en ce qui concerne ces trois segments (desktop, tablet, smartphone)⁶¹.

⁶¹ À titre indicatif, le bounce rate moyen était à cette période de 14% pour Search, 17% pour l'opac KBR, 20% pour Pallas, 34% pour Belgica et 48% pour Warpress.

Voici les données de fréquentation concernant plus précisément ce mois d'avril 2016: SEARCH 169592 visites (desktop : 156780 ; tablet : 8020 ; smartphone : 4180 ; others : 612)

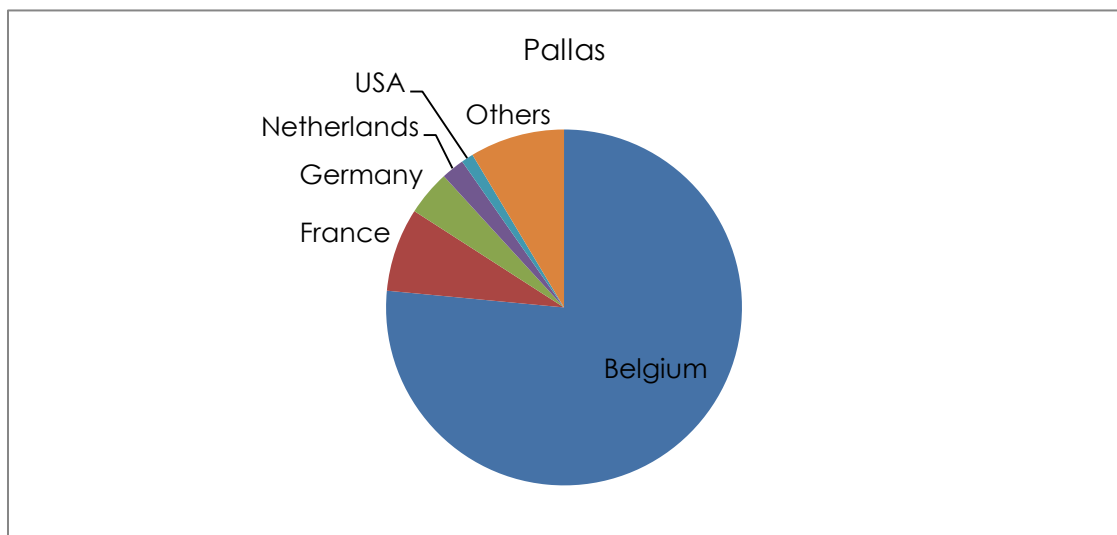
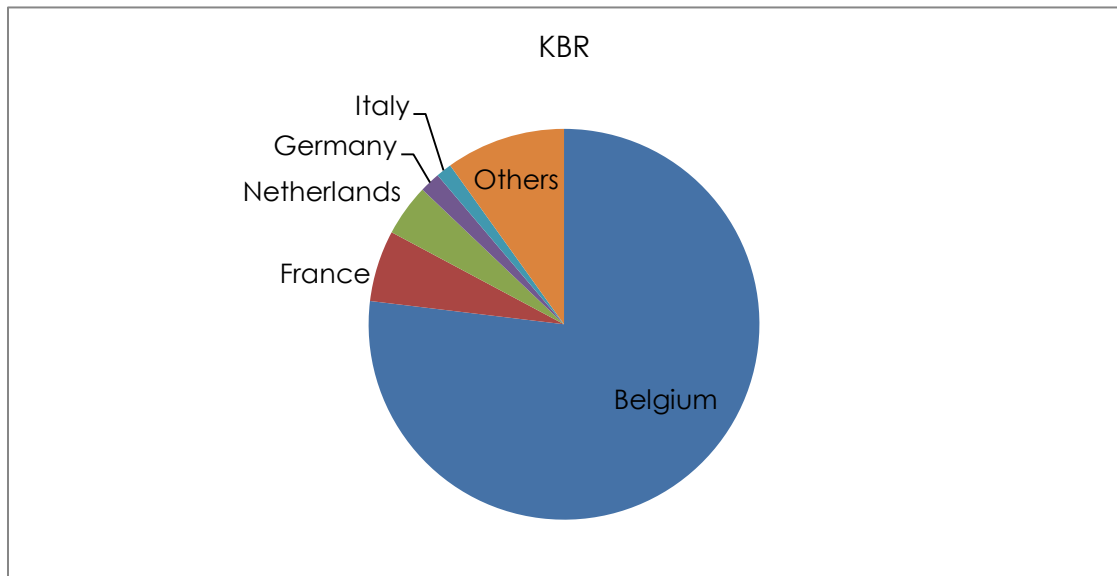
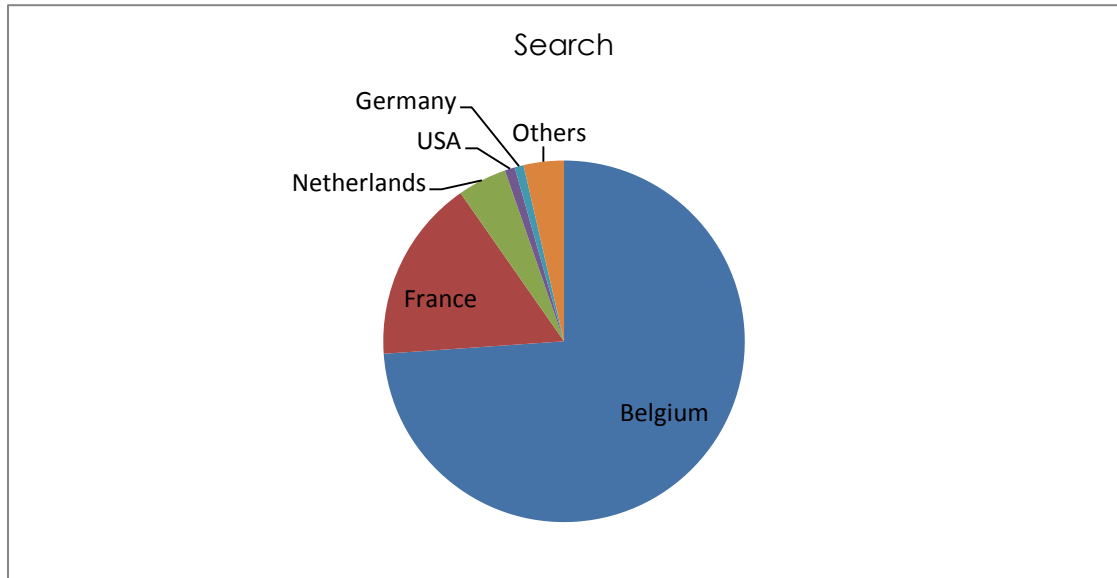
KBR 13576 visites (desktop : 12664 ; tablet : 499 ; smartphone : 373 ; others : 60)

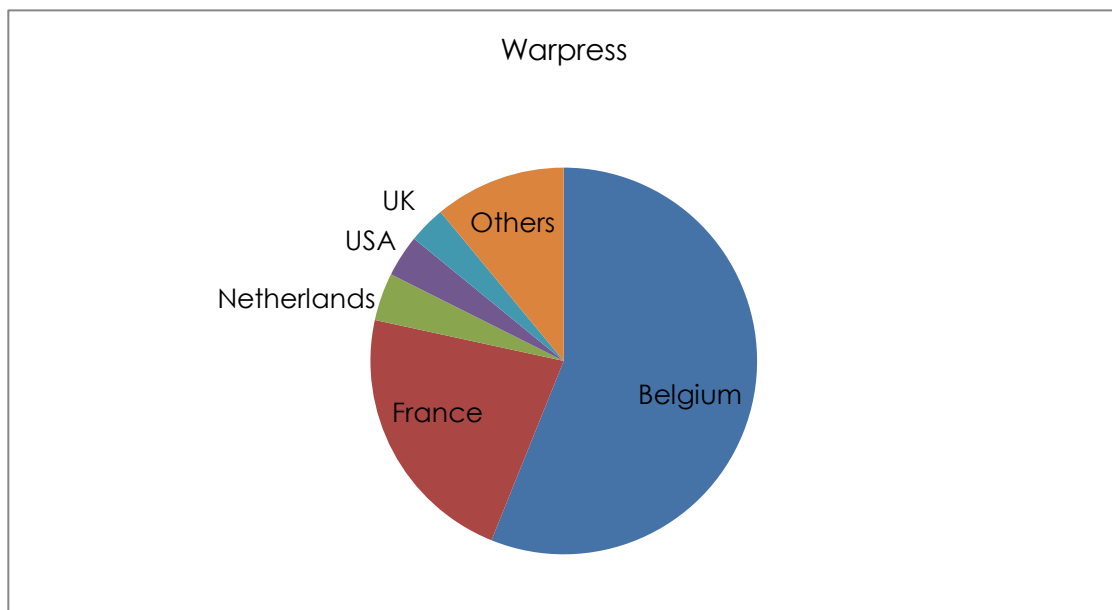
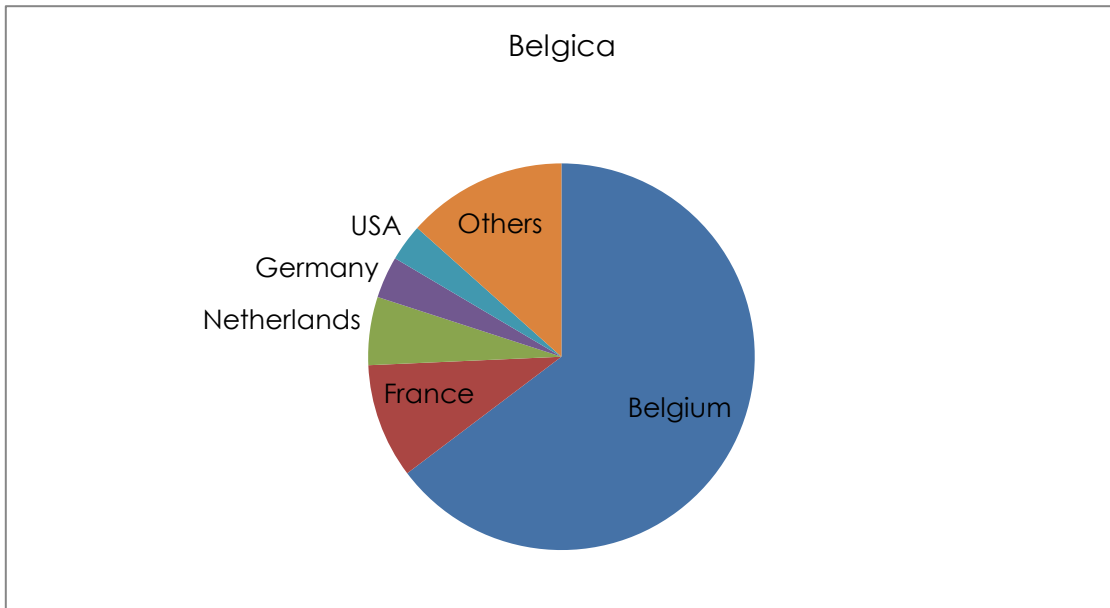
PALLAS 1495 visites (desktop : 1397 ; tablet : 59 ; smartphone : 36 ; others : 3)

BELGICA 4880 visites (desktop : 4430 ; tablet : 272 ; smartphone : 130 ; others : 18)

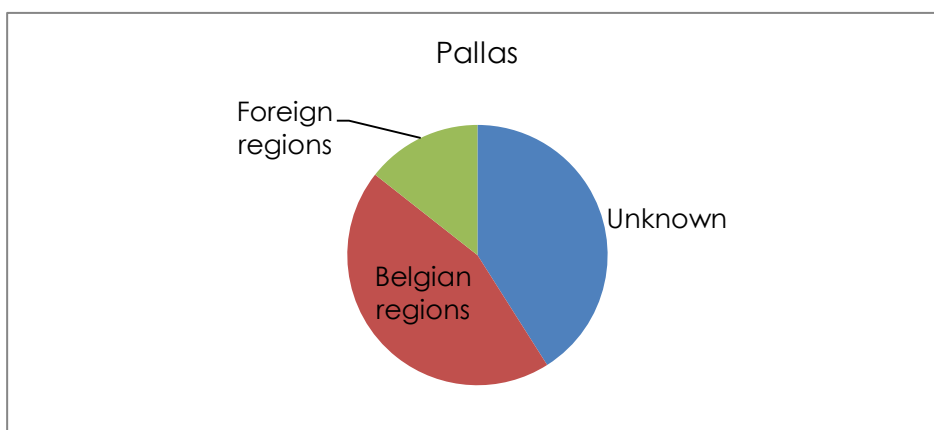
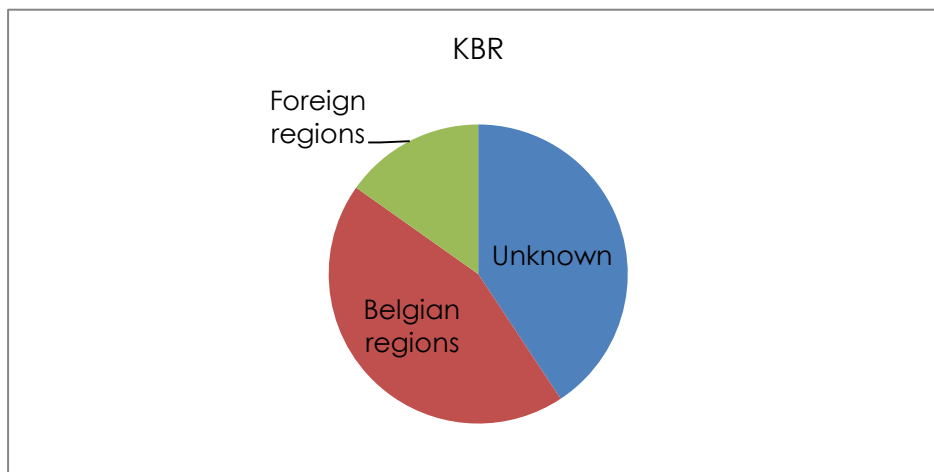
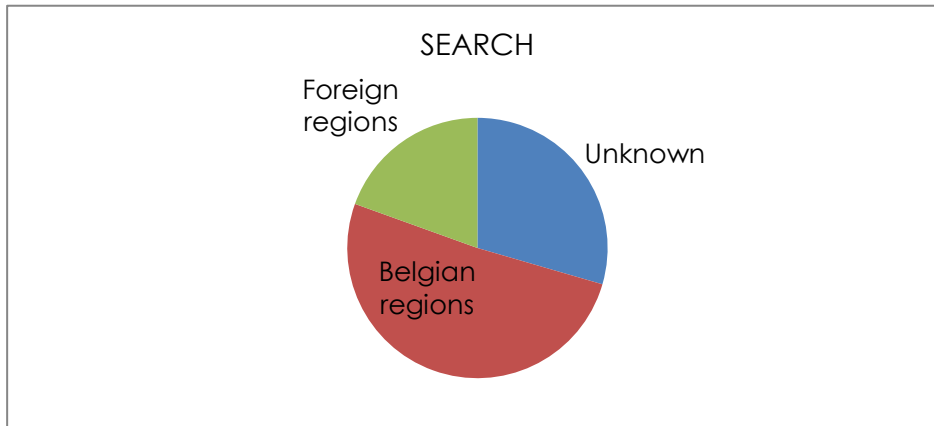
WARPRESS 1849 visites (desktop : 1614 ; tablet : 82 ; smartphone : 122 ; others : 31).

Répartition des visites par pays



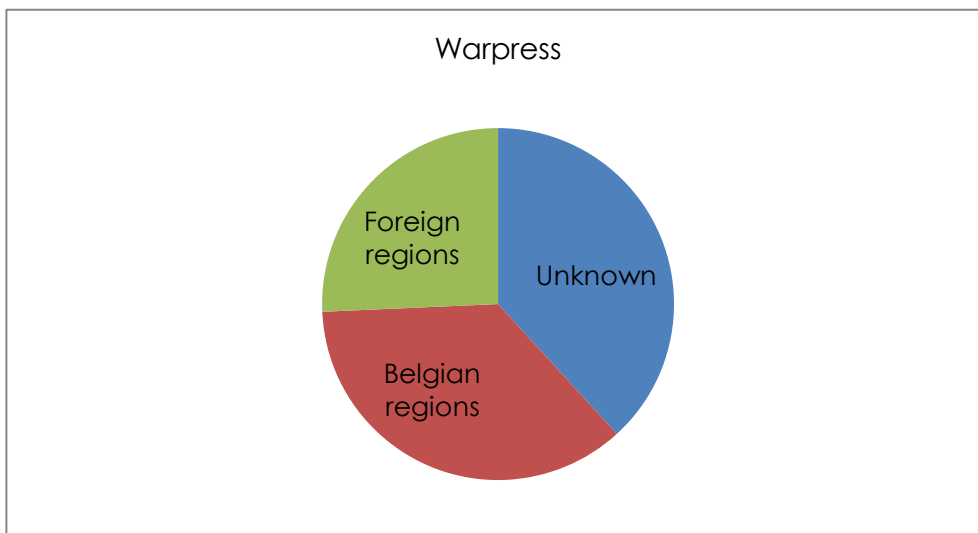
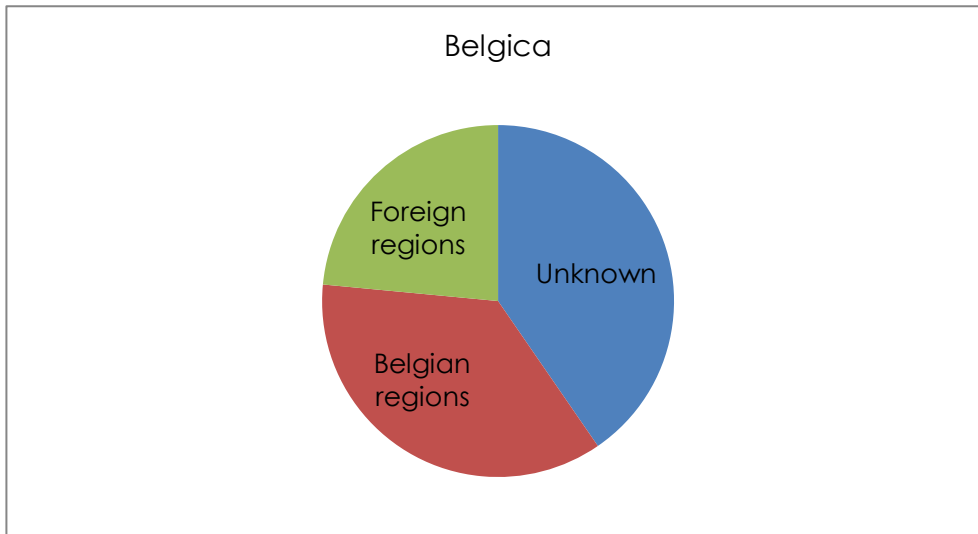


Répartition des visites⁶² par région⁶³

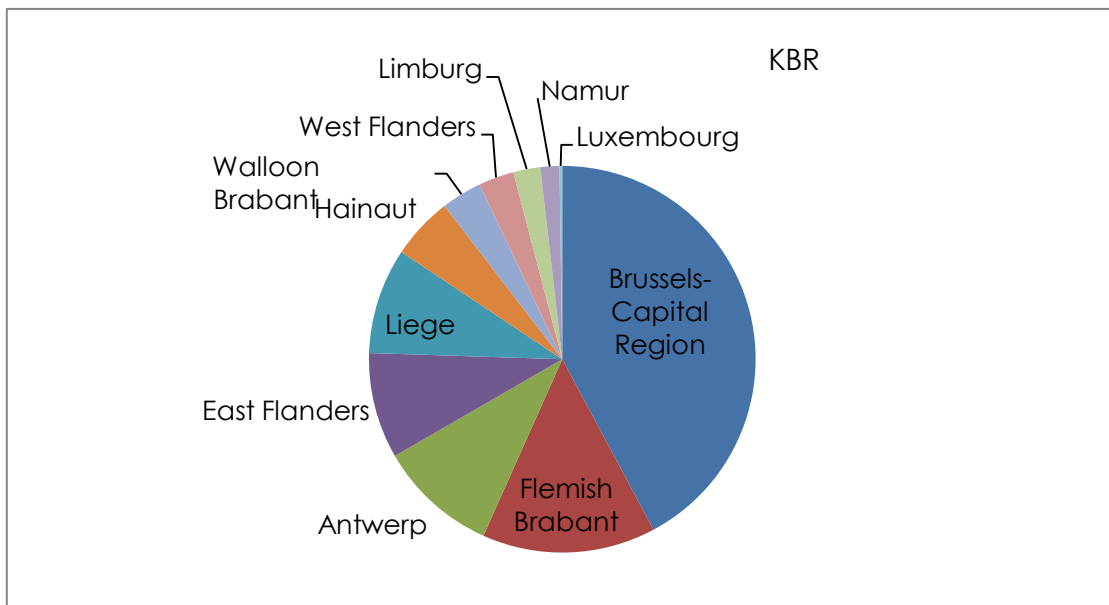
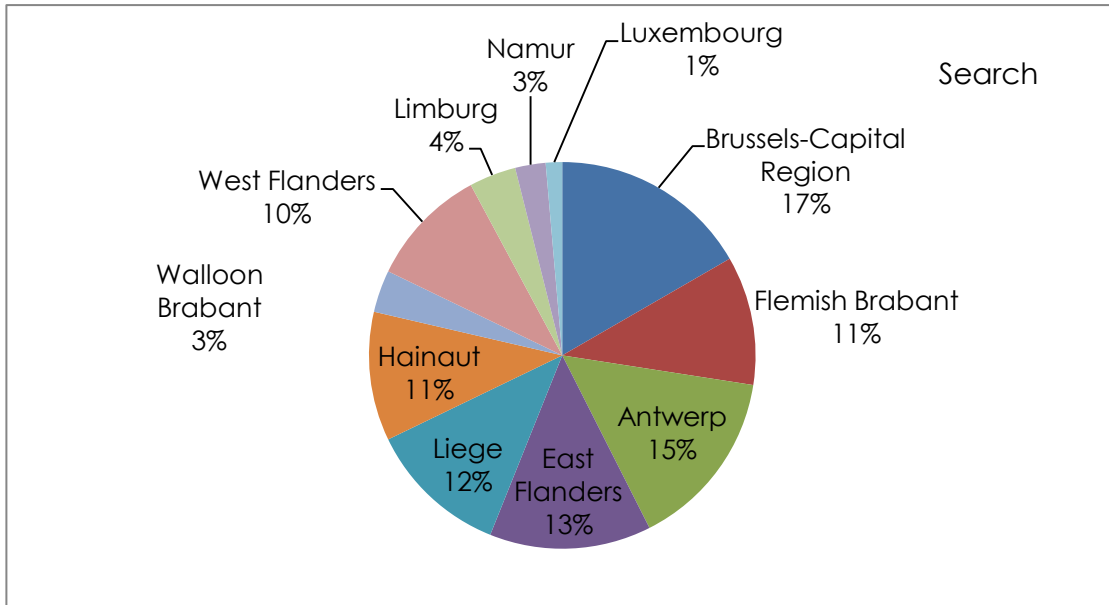


⁶² L'échantillon de données d'analyse pour Search est différent cette fois-ci : il concerne des données allant de janvier à décembre 2016 (raisons techniques).

⁶³ Pour déterminer où une adresse IP est géolocalisée, la base de données GeoLiteCity a été utilisée, tel que recommandé par Piwik. Malheureusement, certains visiteurs ne sont techniquement pas localisables (leur localisation est donc étiquetée comme « unknown »).

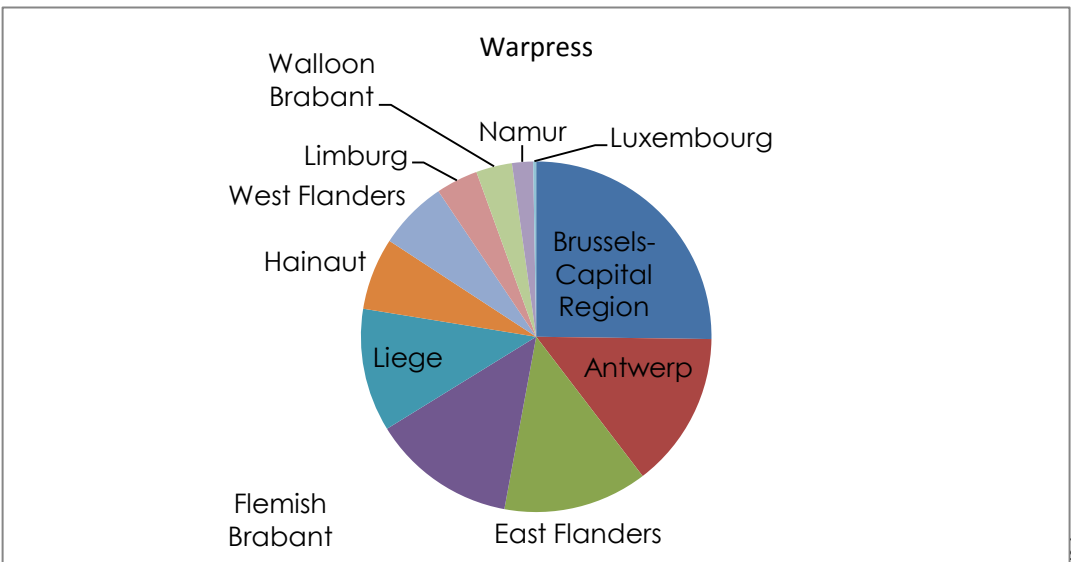
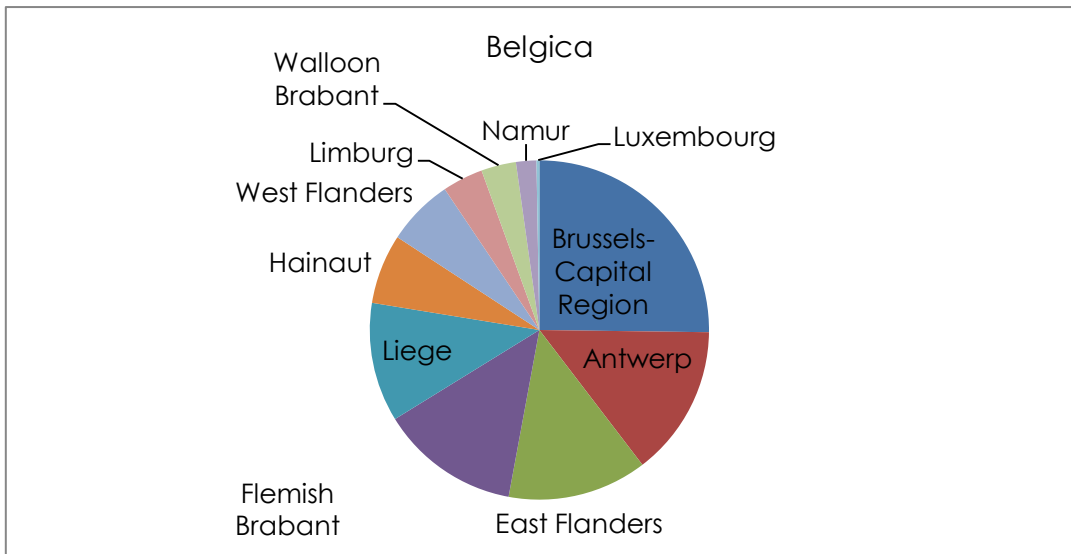
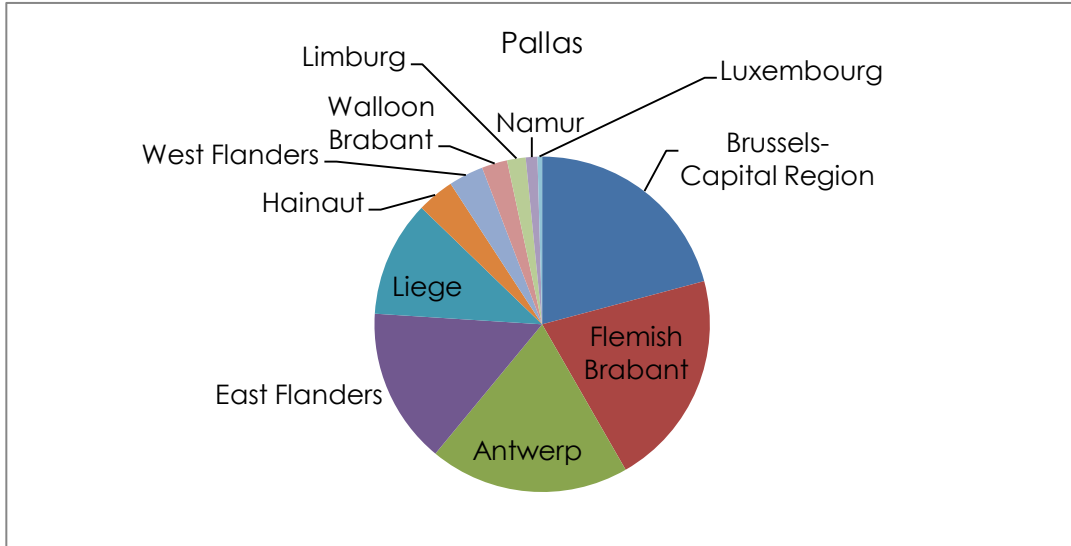


Répartition des visites⁶⁴ issues de Belgique, par région⁶⁵

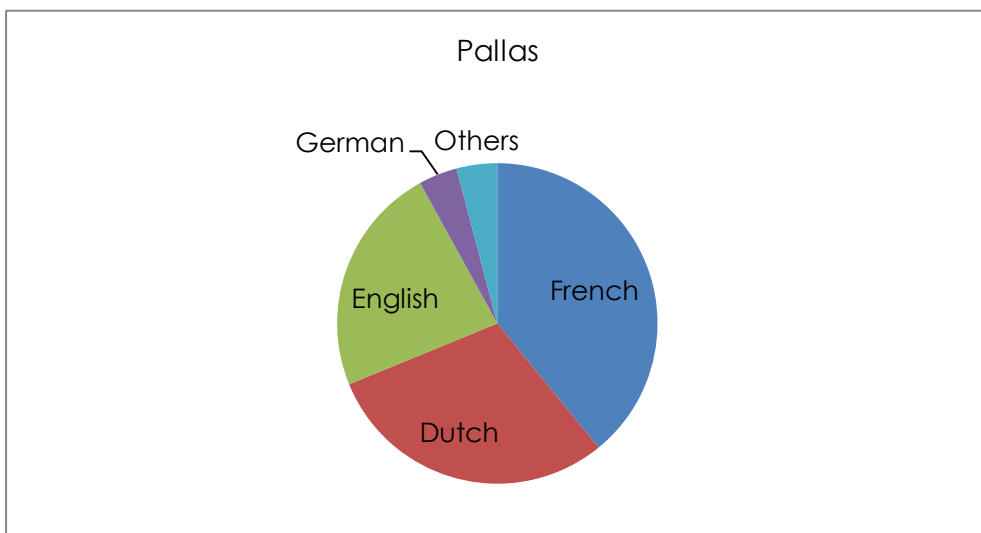
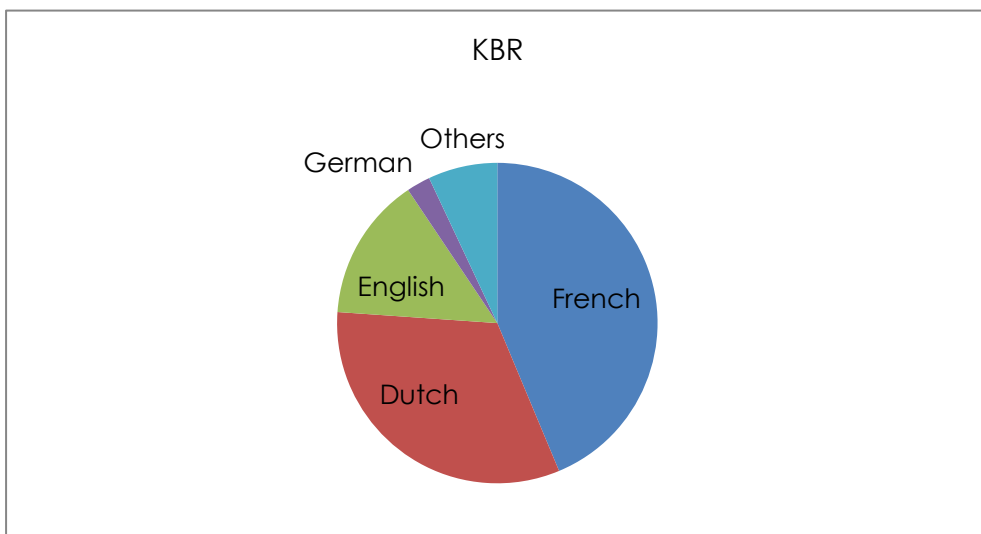
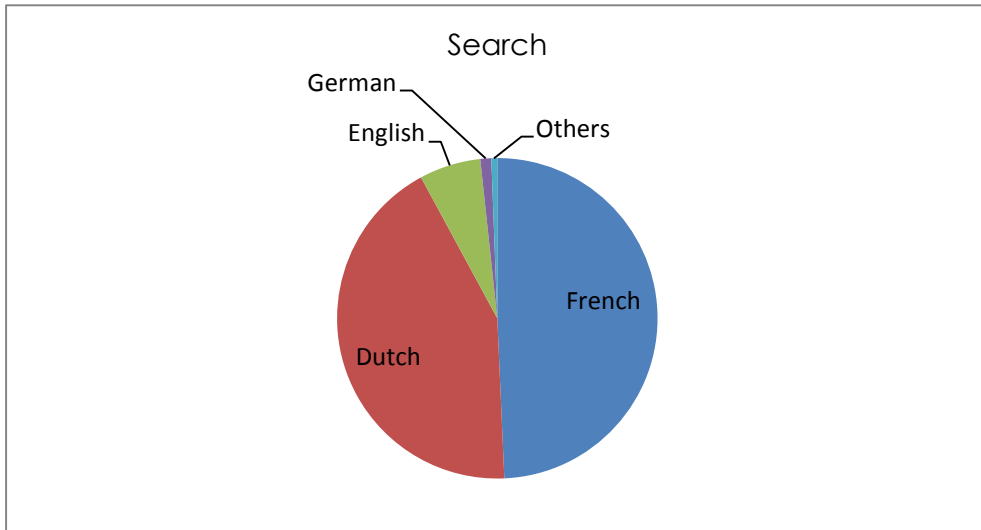


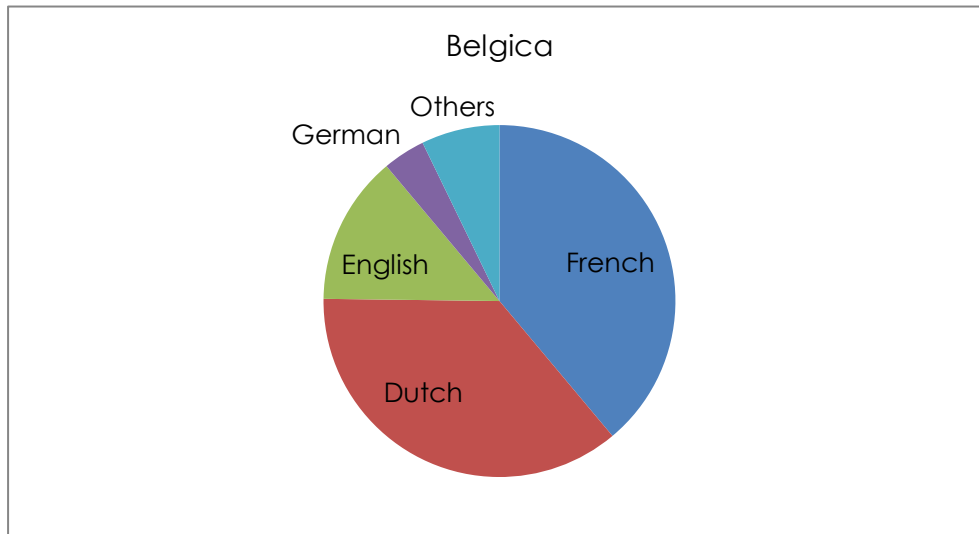
⁶⁴ L'échantillon de données d'analyse pour Search est différent cette fois-ci : il concerne des données allant de janvier à décembre 2016 (raisons techniques).

⁶⁵ Tenir compte des « unknown » montrés dans les graphiques précédents.



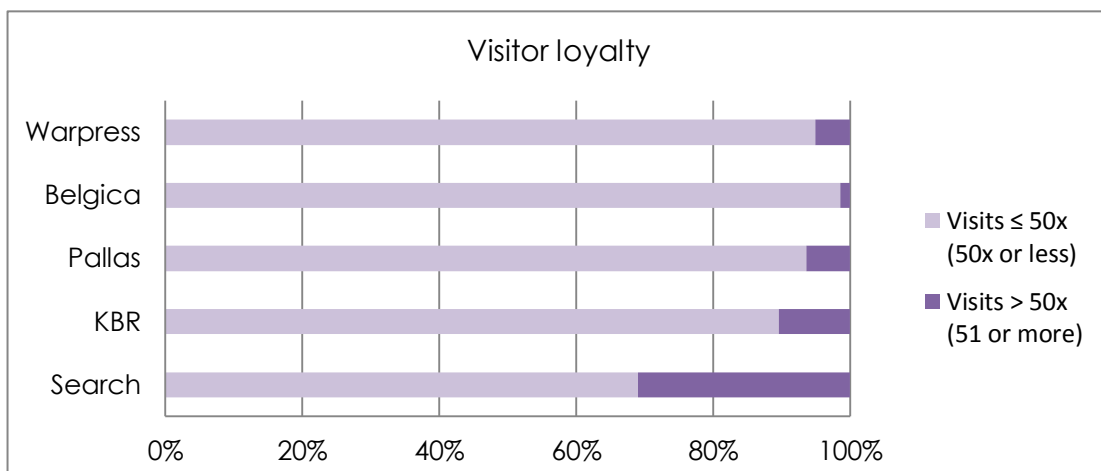
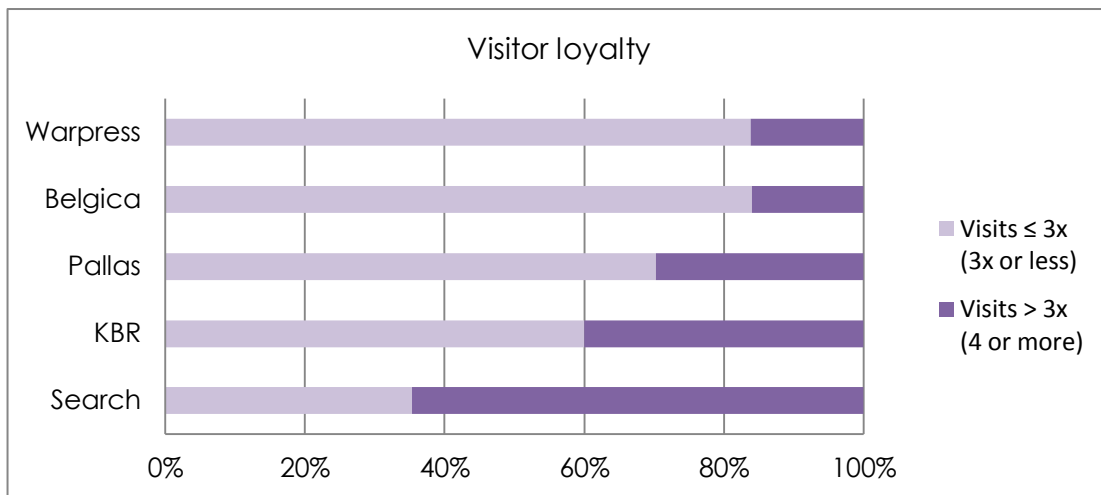
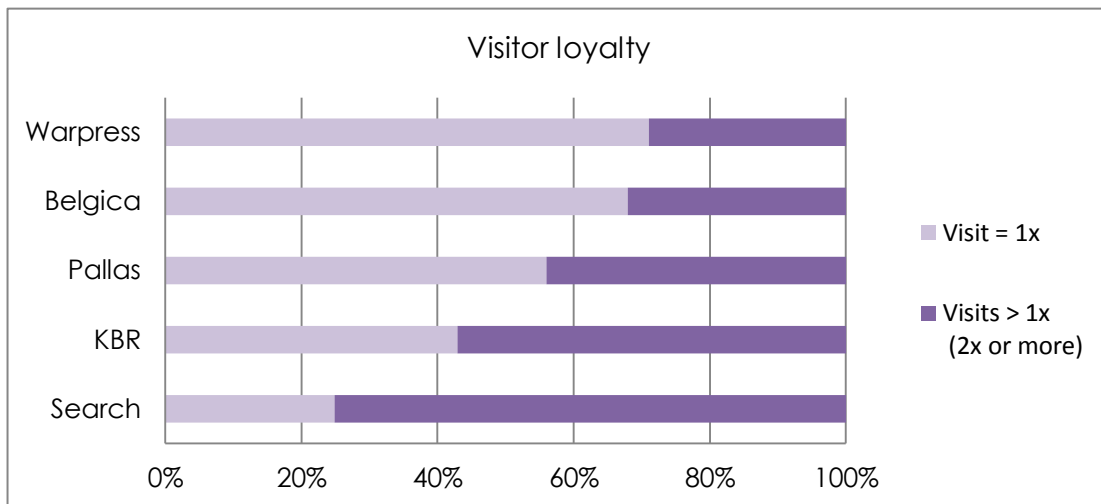
Répartition des visites selon la langue du navigateur web





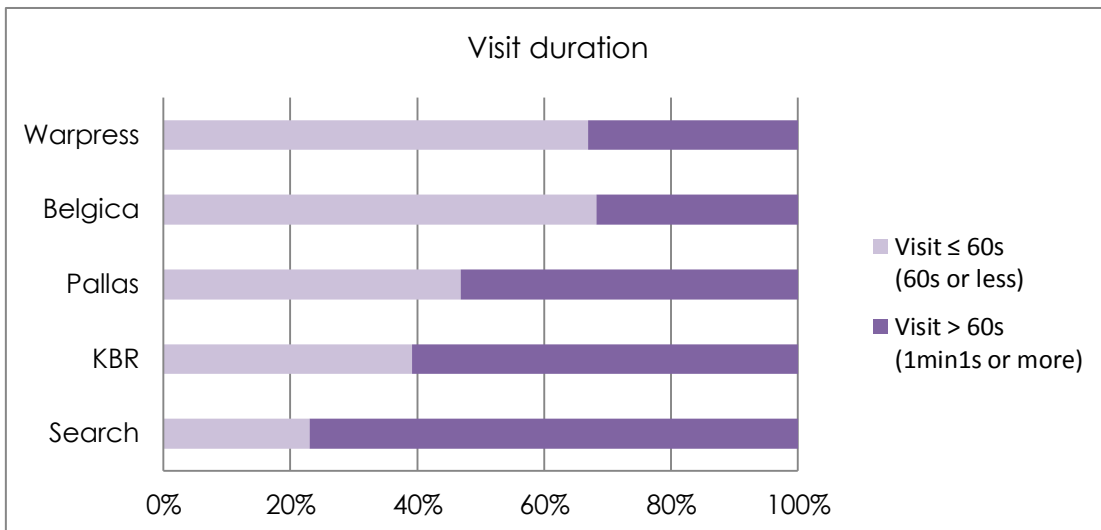
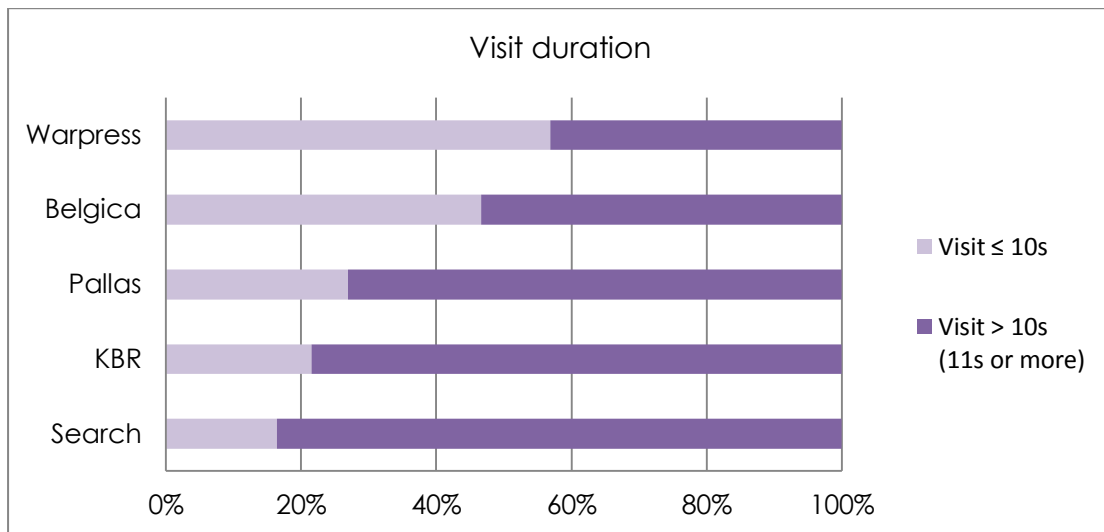
Visitor loyalty

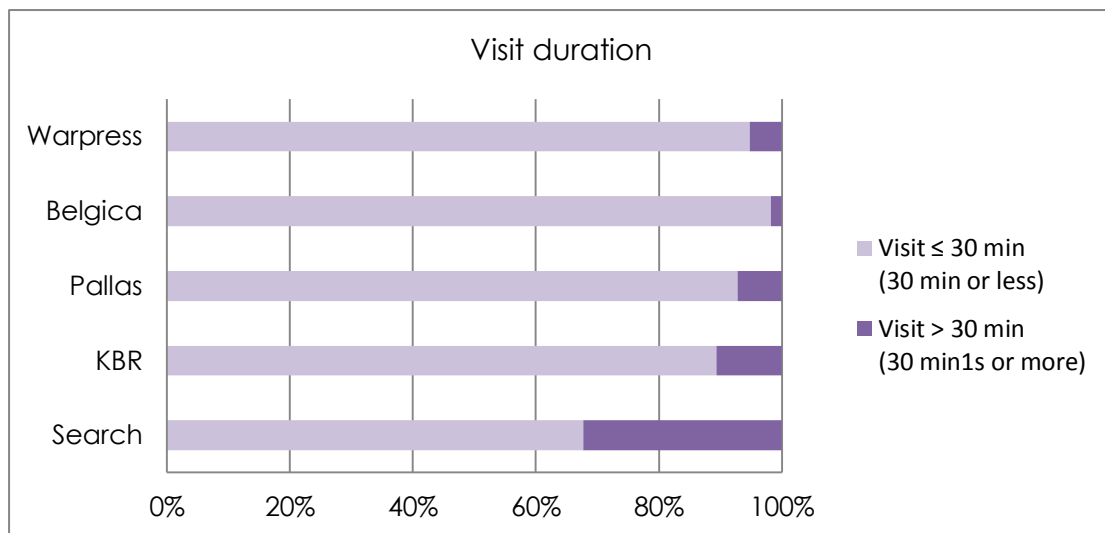
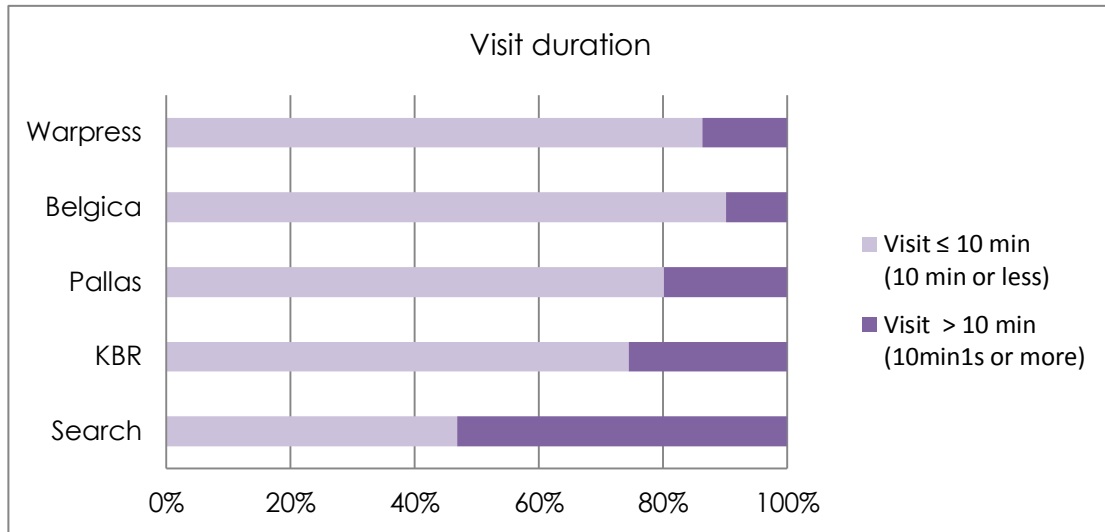
Ce pourcentage permet de savoir si les visiteurs reviennent sur le site. On le calcule grâce au nombre de visites effectuées par une même personne durant la période d'analyse. On peut ainsi savoir, sur tous les visiteurs (100%), combien sont venus plus d'une seule fois ou encore plus de trois fois.



Visit duration

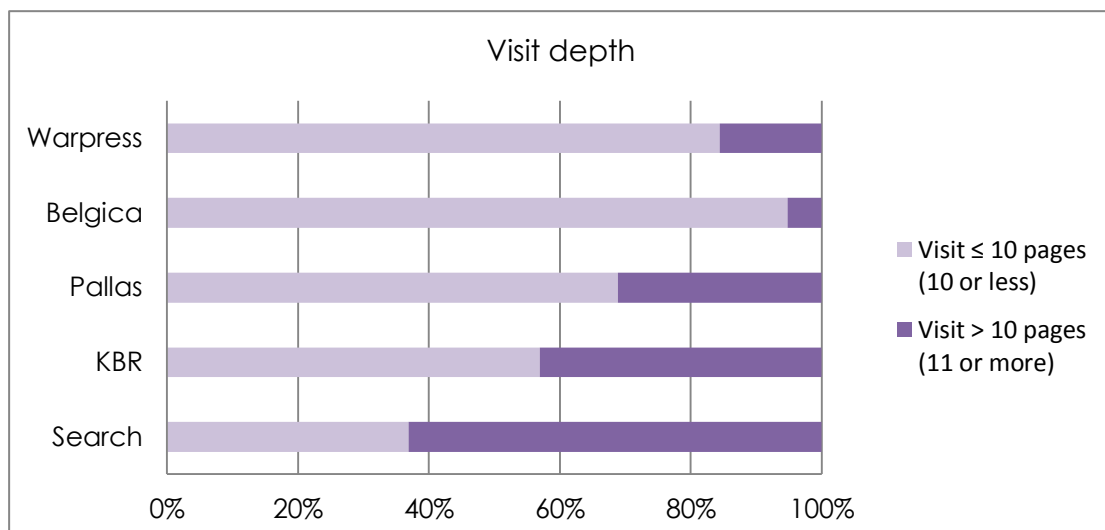
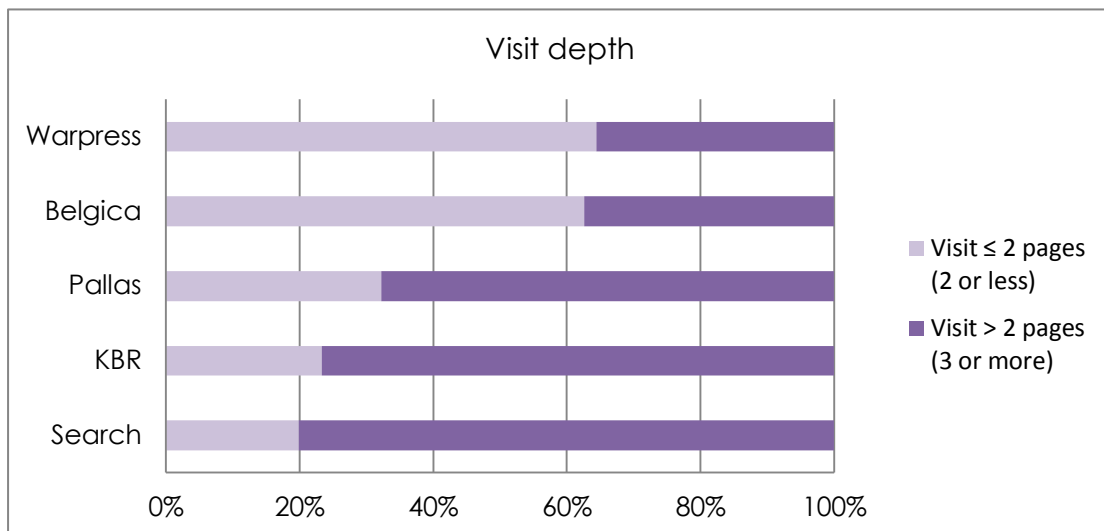
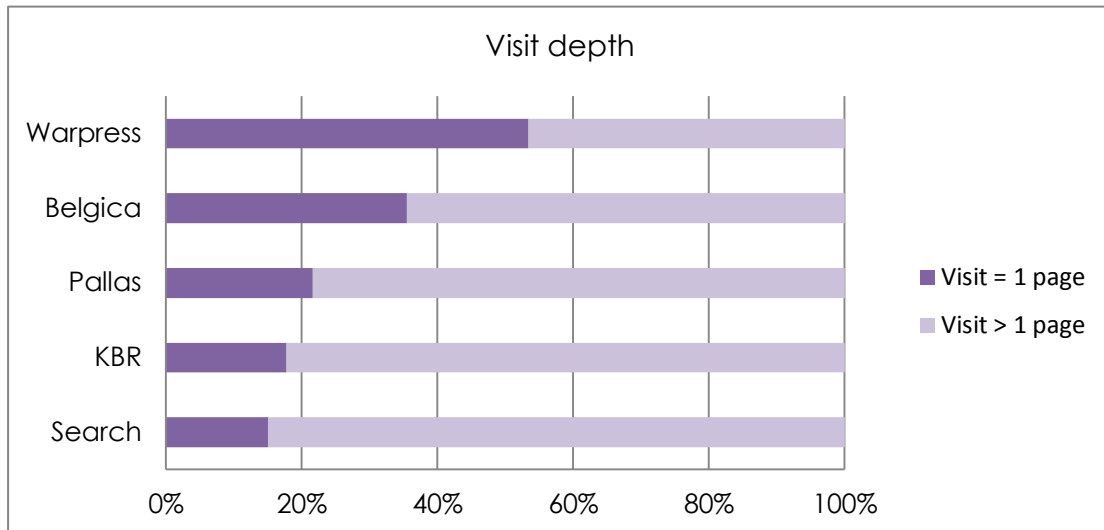
Il s'agit de prendre en compte la durée de chaque visite individuelle, afin d'obtenir des pourcentages globaux. Cela permet d'obtenir des infos plus précises qu'avec la plus classique durée moyenne par visite.





Visit depth

Il s'agit de connaître la profondeur des visites, c'est-à-dire le nombre de pages vues par visite. Cela permet par exemple de voir, ici sous forme de pourcentages, combien de visites comptent plus de trois pages vues.



2.2. Vue comparative : sites de valorisation des collections

Parmi les sites web gérés par les trois institutions participant au projet, nous trouvons trois sites de valorisation (deux expositions virtuelles et le site d'une revue thématique). Si ces derniers ne figurent pas au premier plan de la recherche menée dans le cadre du projet MADDLAIN, ils représentent également des chemins d'accès vers les collections. Nous présentons donc dans cette sous-section un aperçu global de la fréquentation de ces sites. Nous les avons regroupés par commodité, mais nous aimerions toutefois souligner que le site web de la revue thématique a une nature et une fonction tout à fait différentes des sites des deux expositions virtuelles. Il s'agit donc moins de comparer leur succès en termes d'audience, que d'offrir un survol général des habitudes des utilisateurs sur ces sites.

Concrètement, ce sont 2 expositions virtuelles des Archives (Goaaal!⁶⁶, consacré à l'histoire des clubs de football belges, et ExpoCongo⁶⁷, qui met en lumière les traces d'un passé colonial) et du site de la Revue Belge d'Histoire Contemporaine⁶⁸ que publie le CegeSoma.



⁶⁶ <http://goaaal.arch.be>

⁶⁷ <http://www.expocongo.be>

⁶⁸ <http://www.journalbelgianhistory.be>

Accueil | Congo - généralités | Guide des sources en préparation | A propos des Archives de l'État | En savoir plus | Contact | FR - NL - EN

Archives I presume?

Traces d'un passé colonial aux Archives de l'État

Accueil

125^e anniversaire du Traité de Berlin

- Le motif lucratif
 - Exploration, exploitation
 - « Hic sunt dracones » : au tour de la science !
 - Le développement d'un empire colonial
- L'échec du motif lucratif

75 ans de présence belge en Afrique centrale

- La colonie modèle
 - Vie sociale
 - Vie économique
 - Sciences en général
 - Agriculture et agronomie
- L'échec de la colonie modèle

Archives I Presume?

Traces d'un passé colonial aux Archives de l'État

Via cette exposition, les Archives de l'État souhaitent mettre en lumière de nombreux documents témoignant de l'histoire coloniale de la RDC et des territoires sous mandat du Rwanda et du Burundi. Ces archives sont regroupées en deux grands ensembles.

Il y a 125 ans, en février 1885, « l'Association Internationale du Congo » du roi Léopold II fit son apparition sur la scène internationale. La **Conférence de Berlin** annonçait un avenir prometteur pour ce « nouvel État du Congo ». Le terrain était déblayé pour l'exploitation de ce magnifique morceau du gâteau africain. Le colonisateur se forgeait un empire économique et ouvrait ce solisant cœur vierge de l'Afrique à la science occidentale. Mais, quelles ont été les **conséquences pour la population autochtone** ?

Un deuxième volet traite des tentatives de la triade coloniale – l'État, l'Église, les entreprises – pour créer une **colonie modèle**. Les Belges se sont avérés des colonisateurs expérimentés et ont investi dans bon nombre d'initiatives sociales et médicales. Mais, **quels étaient les agendas cachés** ?

Par thème, nous mettons en exergue quelques éléments de l'histoire coloniale, illustrés par des galeries de photographies tirées de nos collections.

Vous souhaitez en savoir plus ? Jetez alors un coup d'œil à nos rubriques « Congo - Généralités », « Sources des Archives de l'État relatives au Congo » ou « À propos des Archives de l'État ».

Des questions ? Nos données figurent sous « Contact ».

Sources: Protocoles et Acte Général de la Conférence de Berlin, 1884-1885, cité dans: STENGERS J., Congo: Mythes et réalité, Paris et Louvain-la-Neuve, Editions Duculot, 1990, p. 86-87; Bruxelles, 2005. Lettre de Léopold II au baron Solvyns du 17 novembre 1877, citée dans: JOYE P. et LEWIN R., Les trusts au Congo, Bruxelles, Société populaire d'éditions, 1961, p. 13.

www.arch.be

NEDERLANDS FRANÇAIS ENGLISH

Search...

QU'EST-CE QUE LA RBHC? | COMMANDES | INFORMATIONS AUX AUTEURS | CONTACT | COMMENT CHERCHER SUR LE SITE ?

ACCUEIL REVUES ARTICLES DÉBAT NOUVELLES



Revue Belge d'Histoire Contemporaine, XLVII, 2017, 1



Les articles de ce premier numéro de 2017 – deux en français et deux en néerlandais – illustrent à nouveau la diversité de la recherche en Histoire contemporaine de la Belgique.

Au menu, des sujets aussi variés que : l'usage de l'humour comme arme rhétorique dans les débats politiques tenus au Parlement flamand entre 2004 et 2014 ; la politique belge de neutralité dans le cadre de la guerre austro-prussienne de 1866 ; la démocratisation de la vie politique après la Première Guerre mondiale et enfin, le rôle joué par la langue dans la construction nationale au sein du tout jeune État belge.

L'équipe de rédaction vous souhaite une bonne découverte !

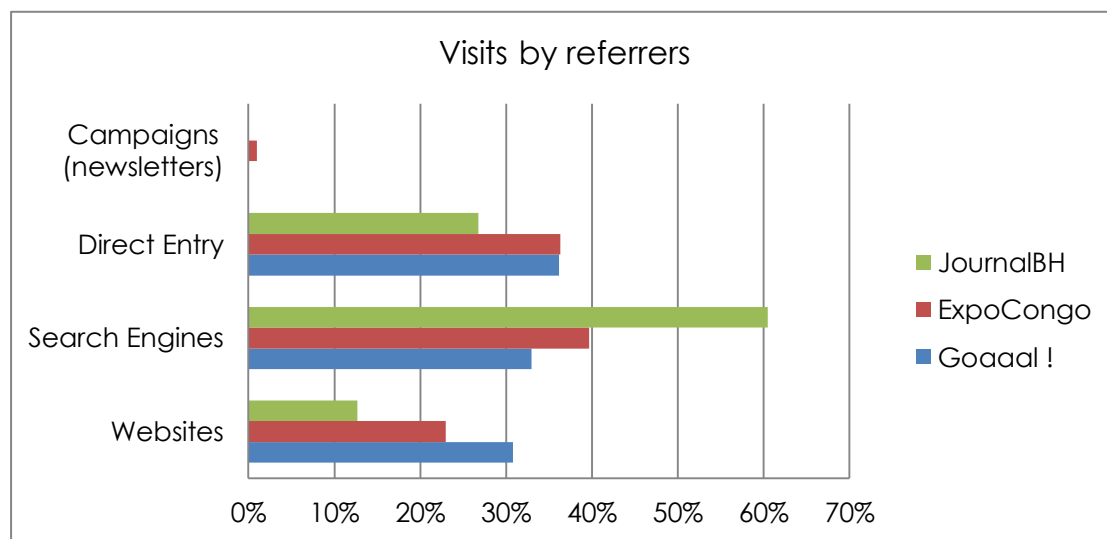
2017 | LIRE LA SUITE >

Le prochain numéro (n° 2-3 / 2017) a du retard

Visites et visiteurs⁶⁹

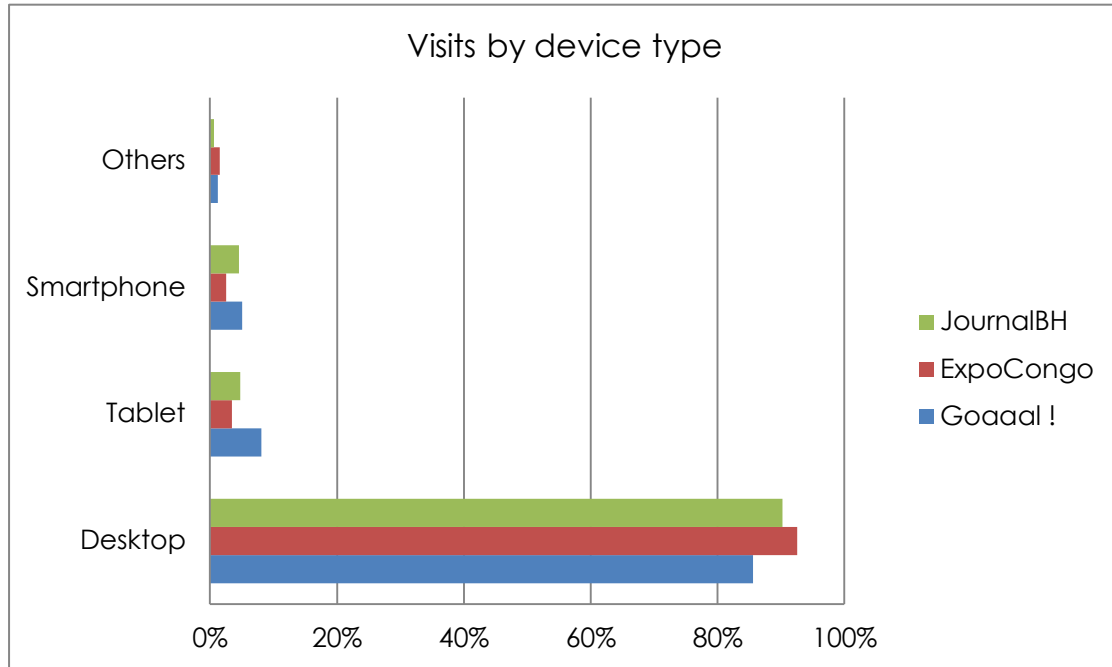
| | Goaaal ! | Expo Congo | Journal BH |
|---------------------------|----------|------------|------------|
| Visites JANVIER | 69 | 64 | 1 482 |
| Visiteurs uniques JANVIER | 62 | 62 | 1 061 |
| Visites FÉVRIER | 28 | 70 | 1 417 |
| Visiteurs uniques FÉVRIER | 25 | 64 | 1 070 |
| Visites MARS | 73 | 18 | 1 398 |
| Visiteurs uniques MARS | 54 | 14 | 1 140 |
| Visites AVRIL | 67 | 12 | 1 406 |
| Visiteurs uniques AVRIL | 47 | 12 | 1 147 |
| Visites MAI | 68 | 25 | 1 357 |
| Visiteurs uniques MAI | 52 | 23 | 1 154 |
| Visites JUIN | 103 | 15 | 988 |
| Visiteurs uniques JUIN | 93 | 12 | 844 |

Répartition des visites par type d'entrée sur le site

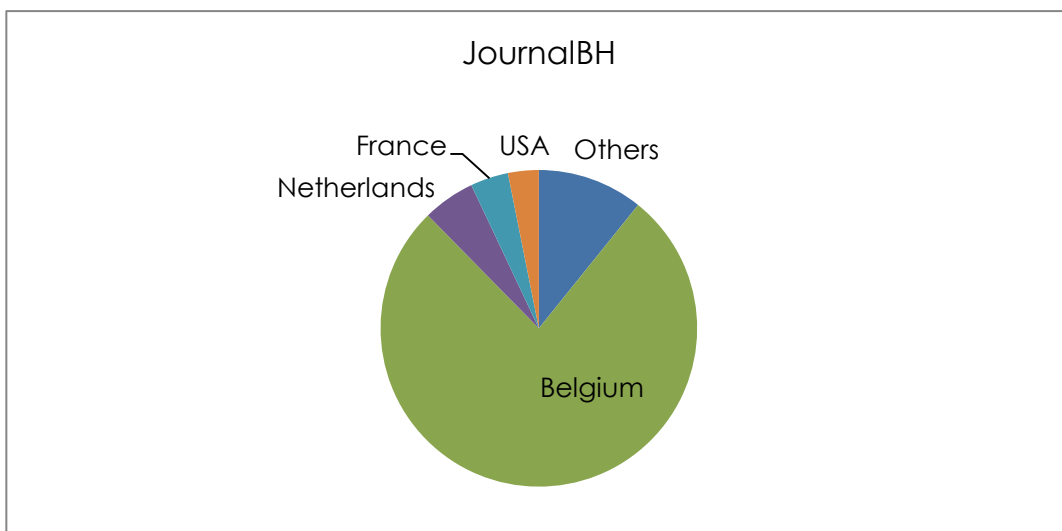
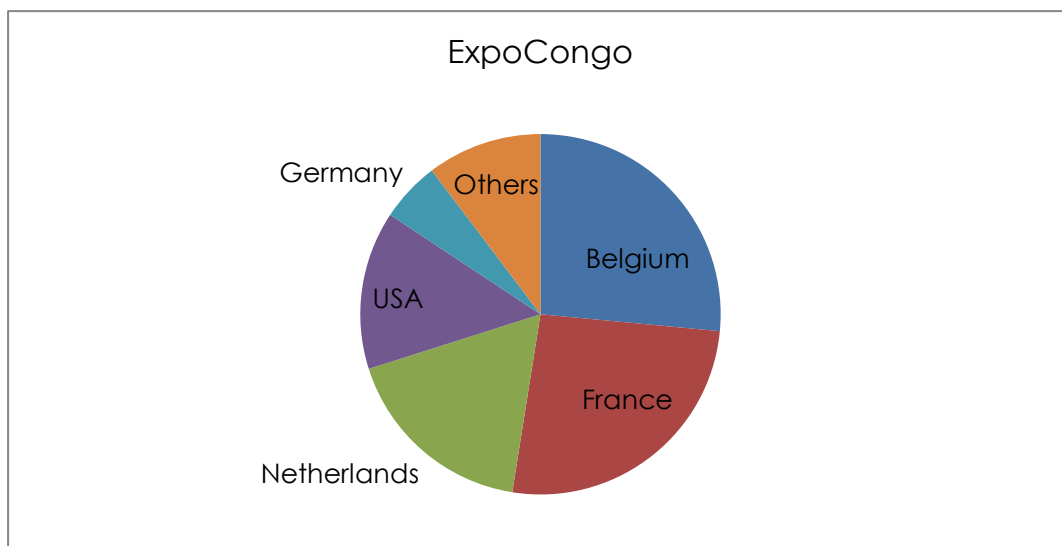
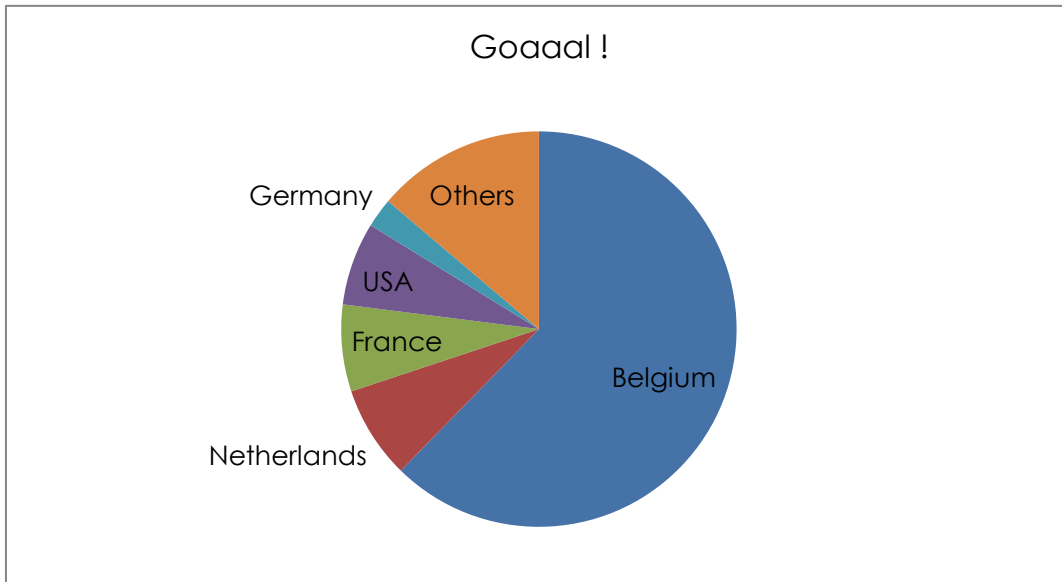


⁶⁹ Dans un souci de concision, les descriptions de métriques proposées dans la sous-section précédente n'ont pas été reprises ici.

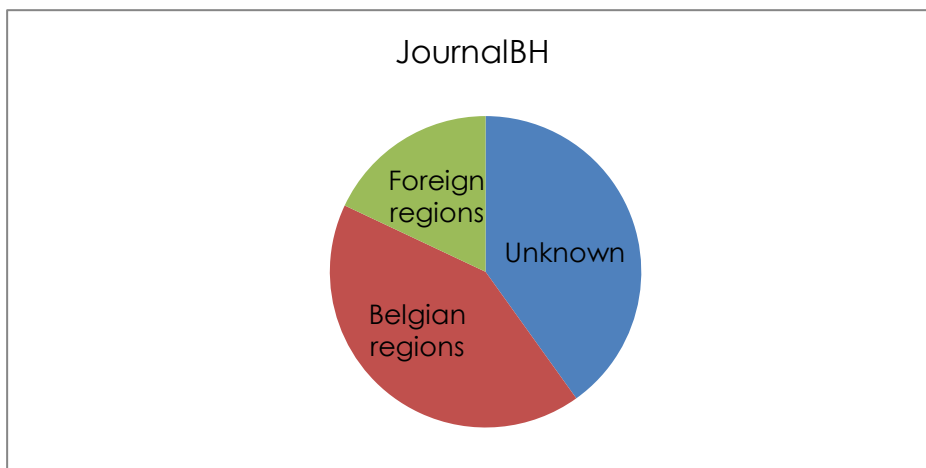
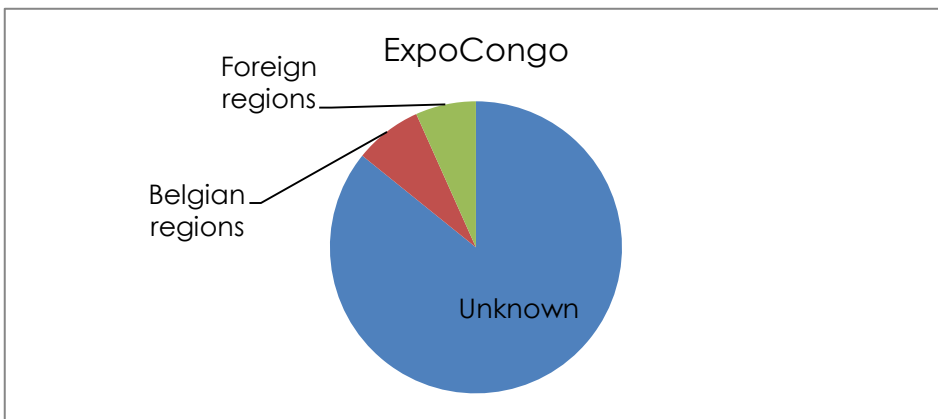
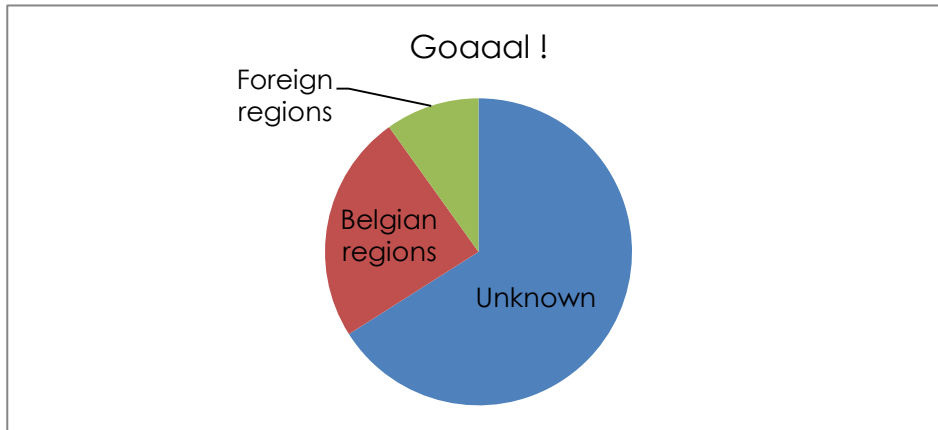
Répartition des visites par type de support



Répartition des visites par pays

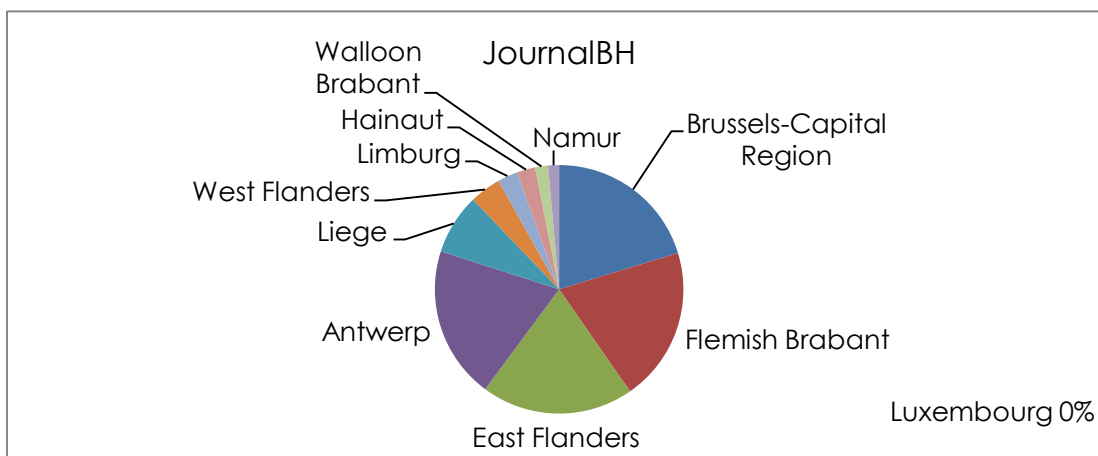
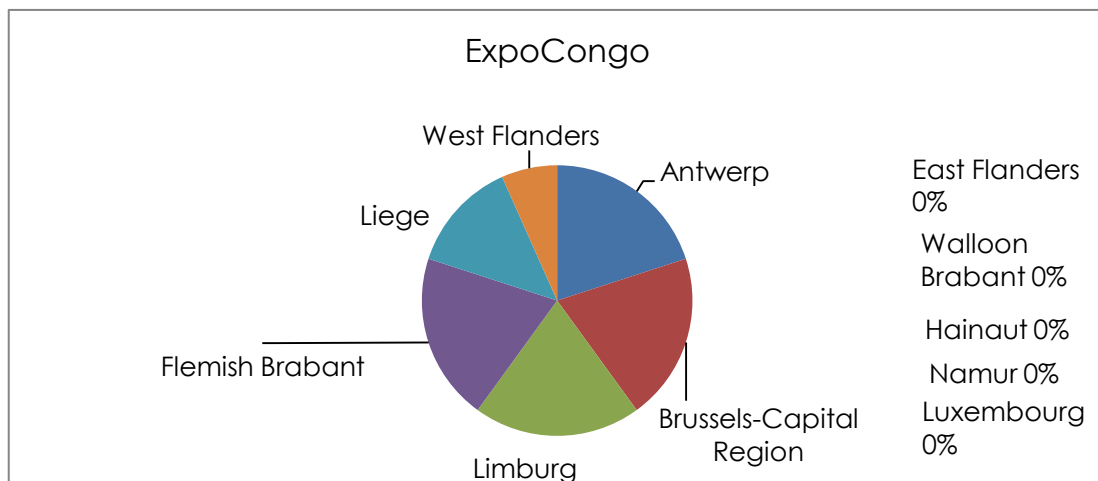
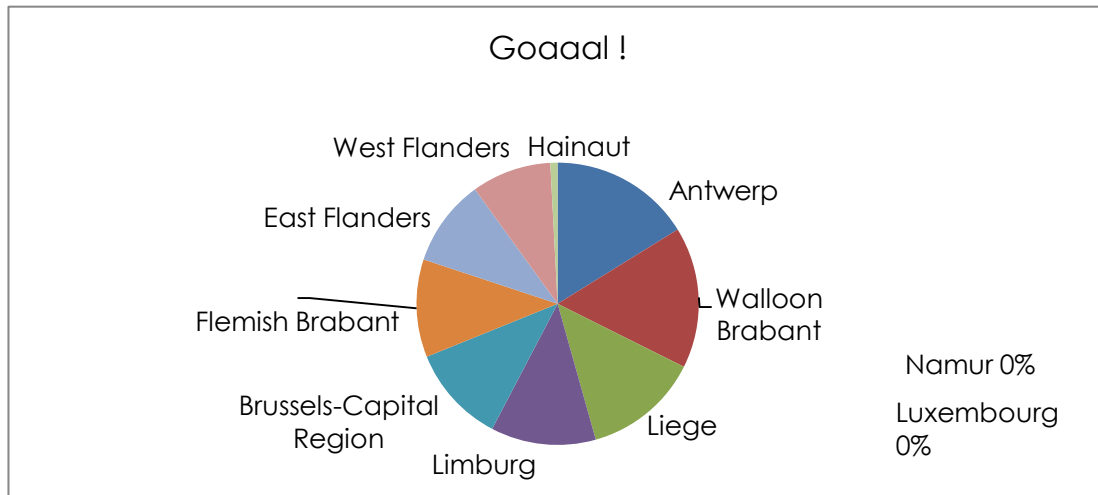


Répartition des visites par région⁷⁰



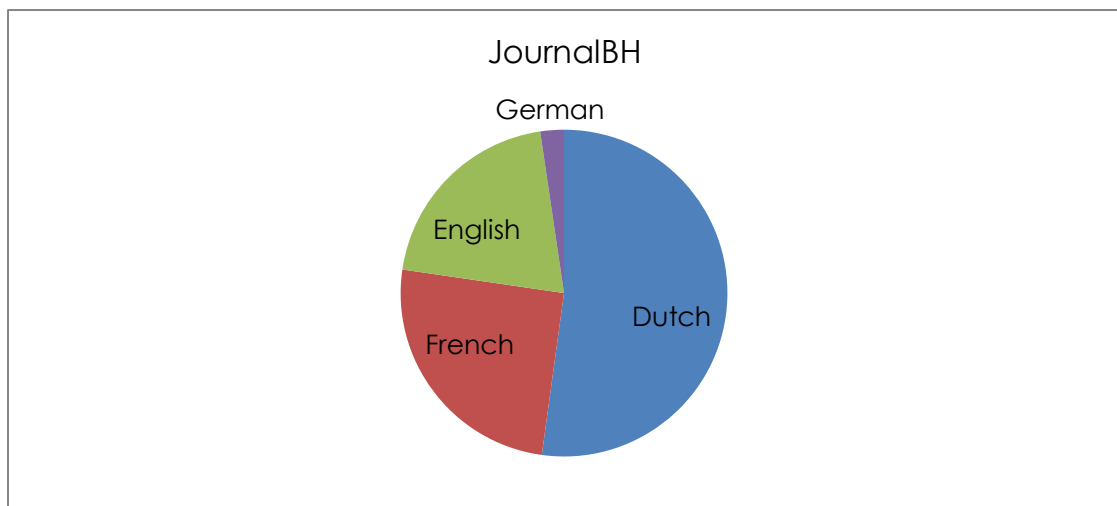
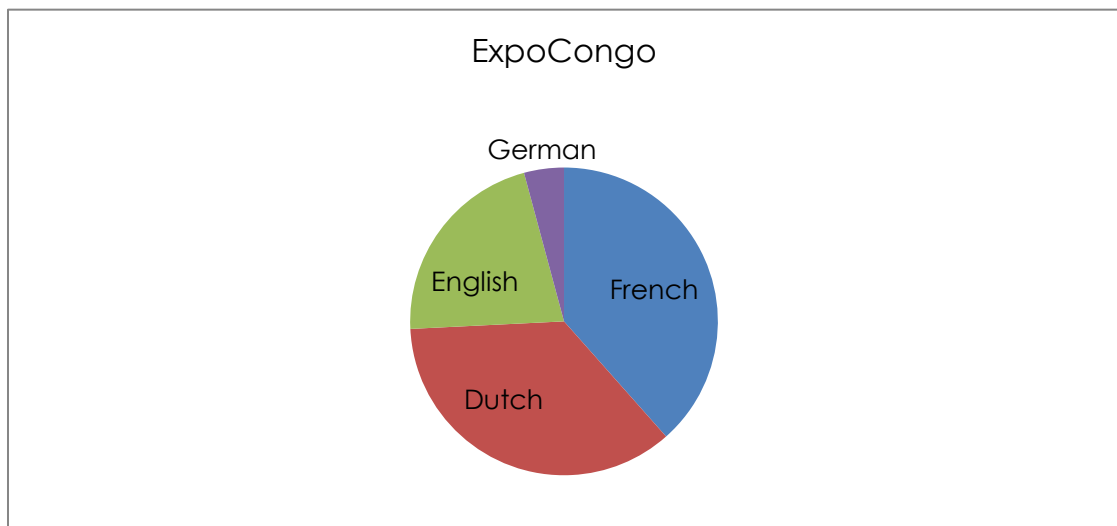
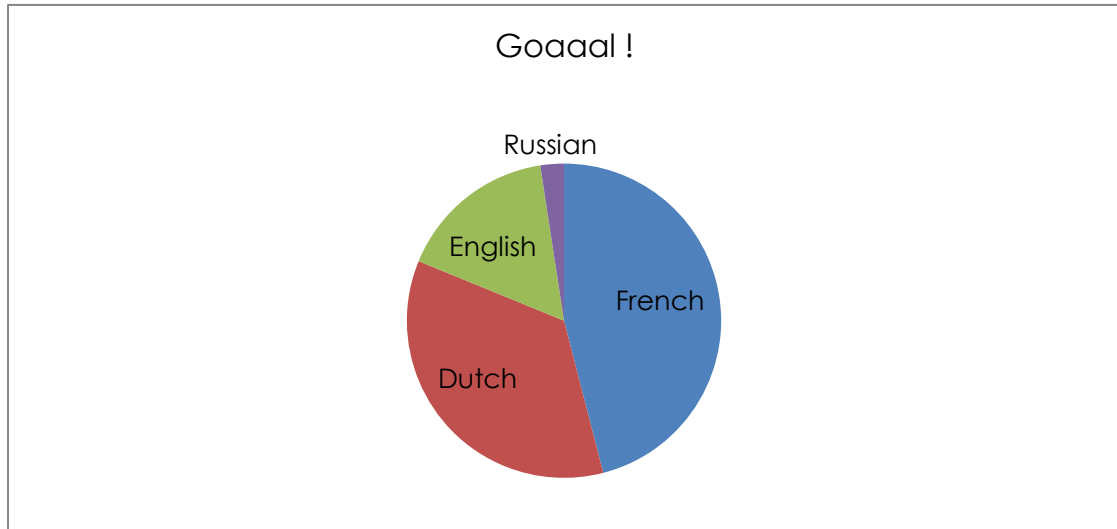
⁷⁰ Pour déterminer où une adresse IP est située géographiquement (géolocalisation), la base de données free GeoLiteCity a été utilisée, tel que recommandé par Piwik. Malheureusement, certains visiteurs ne peuvent pas être localisés (« unknown » label).

Répartition des visites issues de Belgique, par région ⁷¹

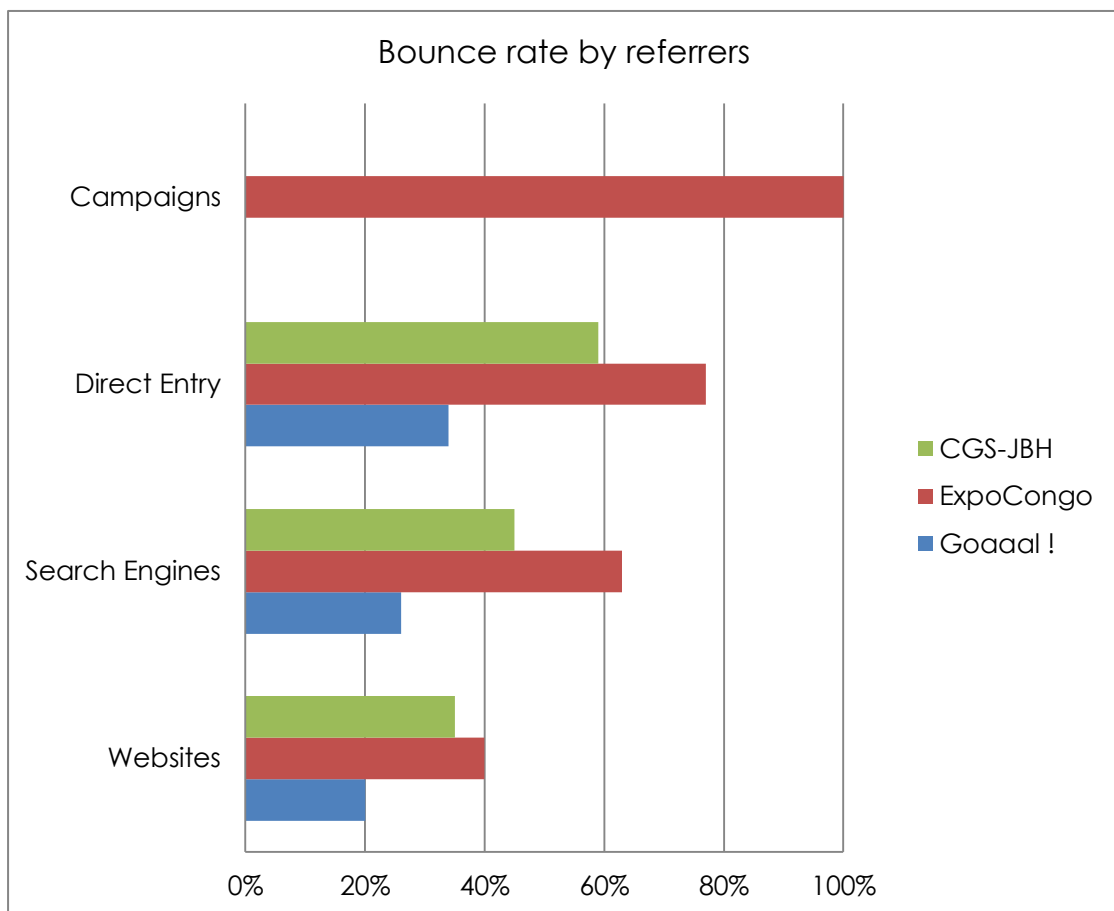
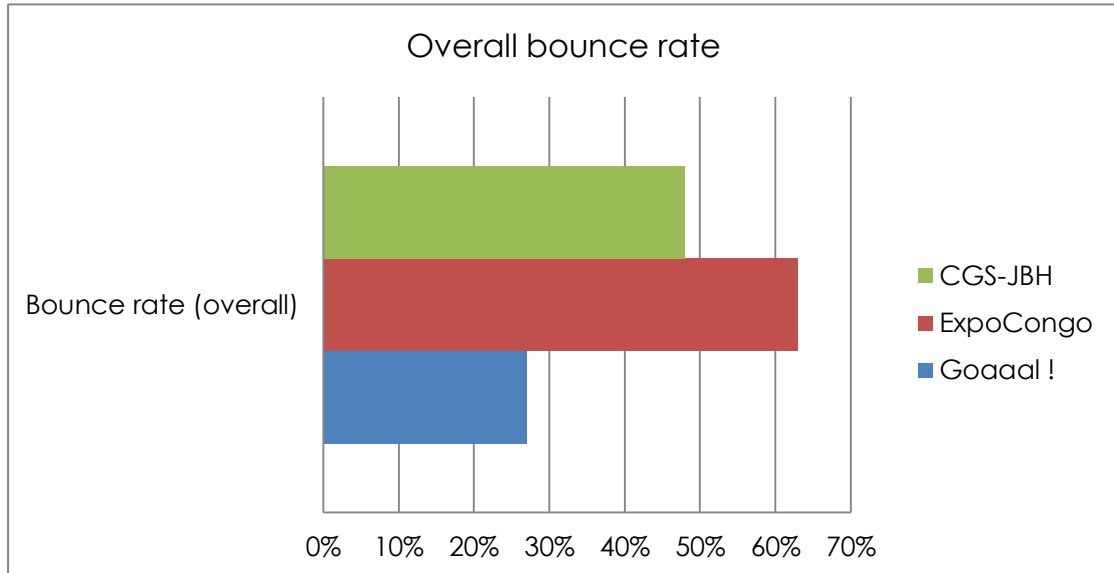


⁷¹ Il faut bien tenir compte des régions non identifiées (« unknown ») qui sont visibles dans les graphiques précédents.

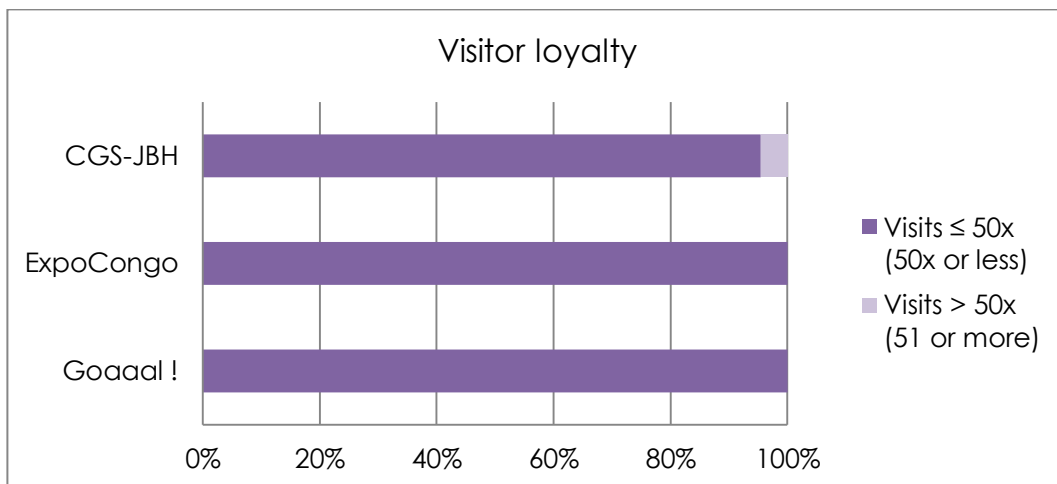
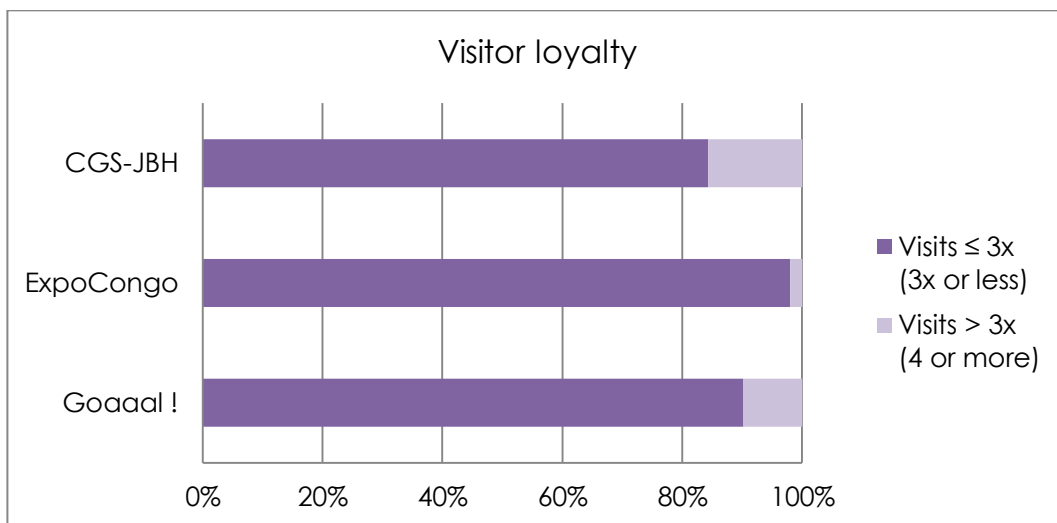
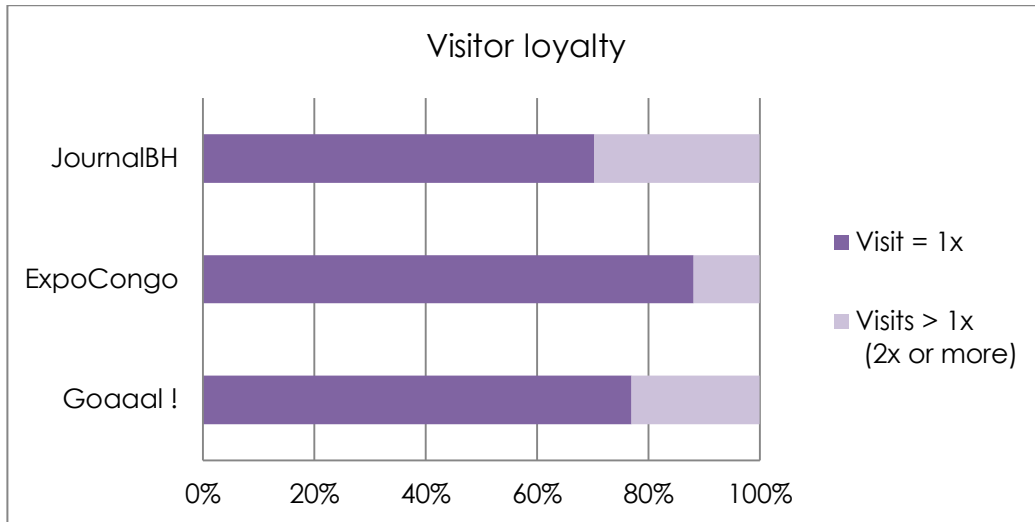
Répartition des visites selon la langue du navigateur web



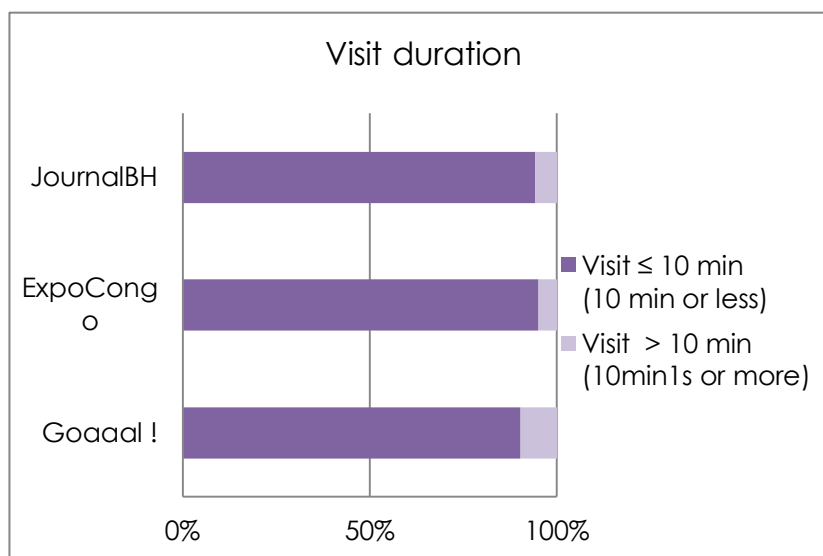
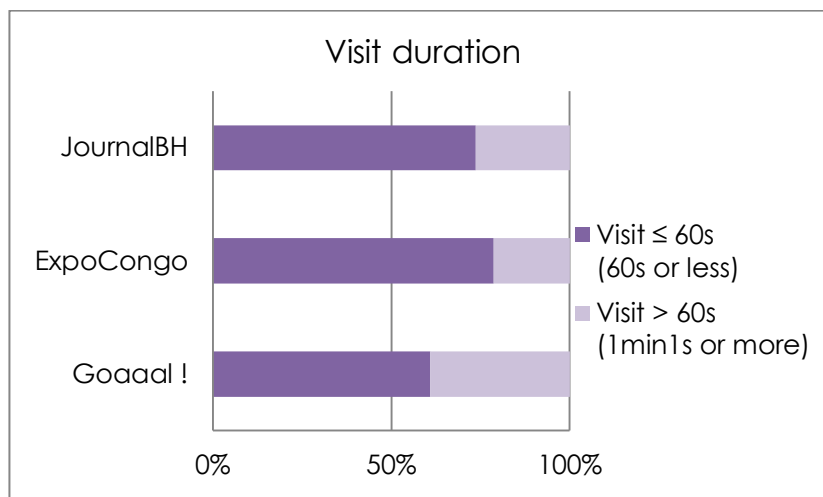
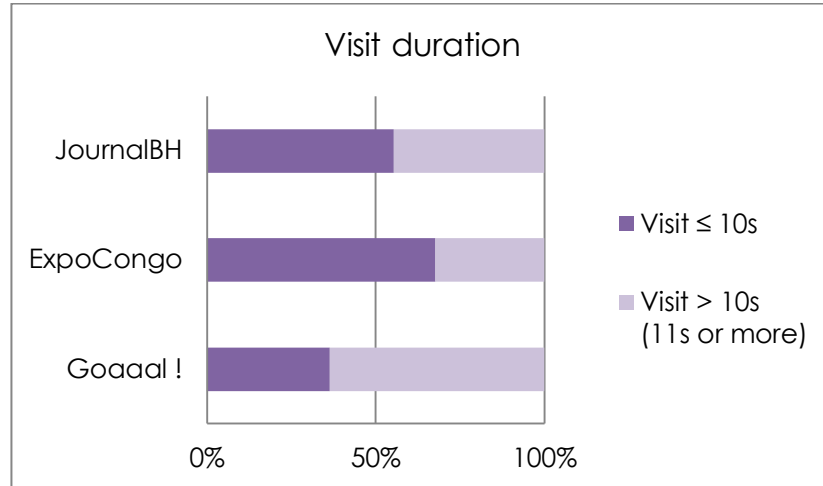
Bounce rate



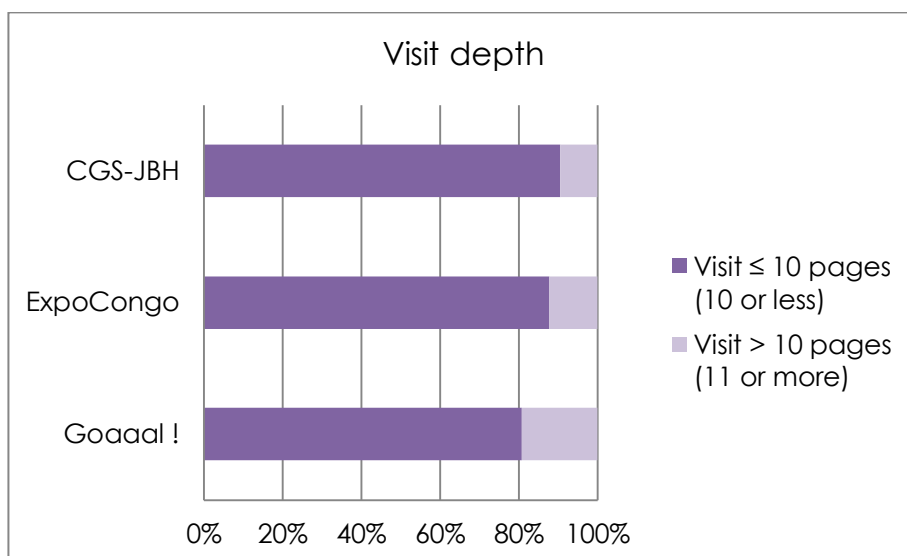
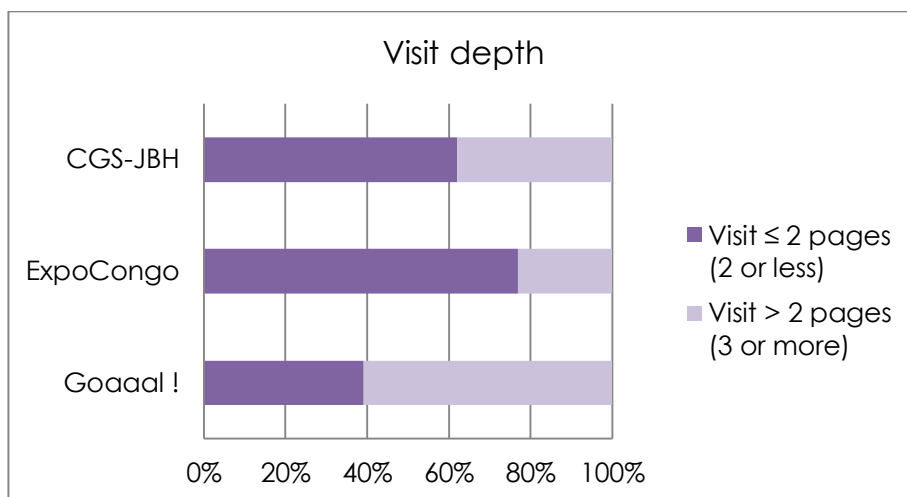
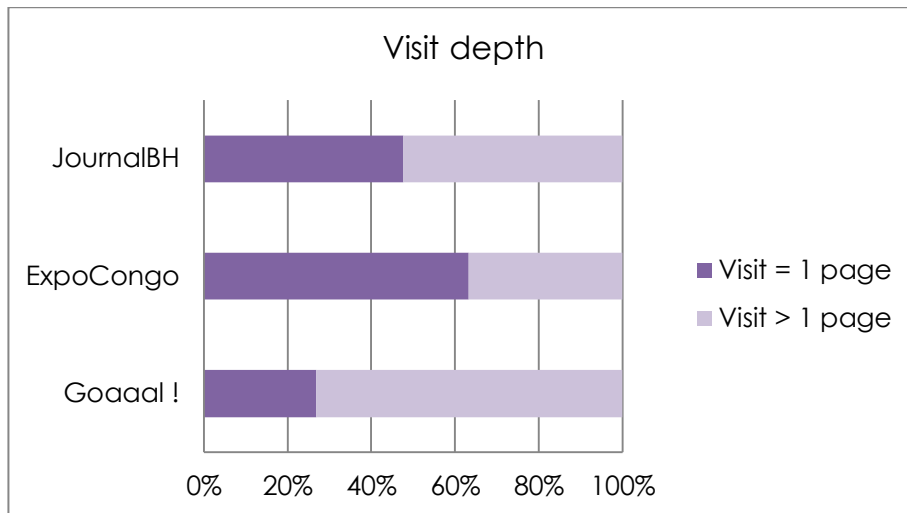
Visitor loyalty



Visit duration



Visit depth



2.3. Analyses spécifiques : sites des catalogues en ligne et sites d'accès aux collections numérisées

Cette partie du rapport consiste en une analyse approfondie des catalogues en ligne principaux de chaque institution. En ce qui concerne le nombre de visites, le profil des visiteurs et leurs habitudes de visites, la section précédente présentait déjà un certain nombre d'informations. Nous les reprenons ici de façon synthétique et les complétons à l'aide de données supplémentaires. Les différents sites sont présentés selon la répartition institutionnelles au démarrage du projet MADDLAIN (AGR, CegeSoma, KBR).

Enfin, veuillez noter que si aucune information n'est donnée concernant la date, il s'agit toujours du même échantillon d'analyse couvrant une période de 6 mois (01.01.2016-01.07.2016).

2.3.1. Analyse spécifique : SEARCH

The screenshot shows the website interface for 'La recherche aux Archives de l'État'. At the top, there is a navigation menu with links for Home, Archives, Producteurs, Personnes, Bibliothèque, and Thèmes. Below the menu is a search bar with a 'Recherche' button and a login section with fields for 'Identifiant' and 'Mot de passe'. The main content area is titled 'La recherche aux Archives de l'État' and contains two update notices: one about a new collection of cadastral plans and documents, and another about parish registers and civil registers. It also includes a link to a video on how to use the search engines.



COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 800 000 reprises.

Il est important de noter que parmi ces 800 000 visites, plus d'une visite sur dix s'est terminée après que le visiteur n'ait vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 15% pour le catalogue Search.

Par ailleurs, il nous est possible de savoir que plus de trois visiteurs sur quatre sont venus plus d'une fois sur le site (75,1%), plus de 60% sont venus quatre fois ou plus (64,7%) et près d'un tiers des visiteurs sont venus plus de 50 fois (31%).

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que près de trois visites sur quatre (73,9%) sont effectuées depuis la Belgique, qui est suivie par la France (16,4%) et les Pays-Bas (4,4%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 31,3% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (8,1%), Anvers (7,6%), le Brabant flamand (6,8%), la Flandre orientale (6,7%) et Liège (6,0%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (49,3%), suivis de près par le néerlandais (42,8%) ; viennent ensuite l'anglais (6,2%) et l'allemand (1,1%).



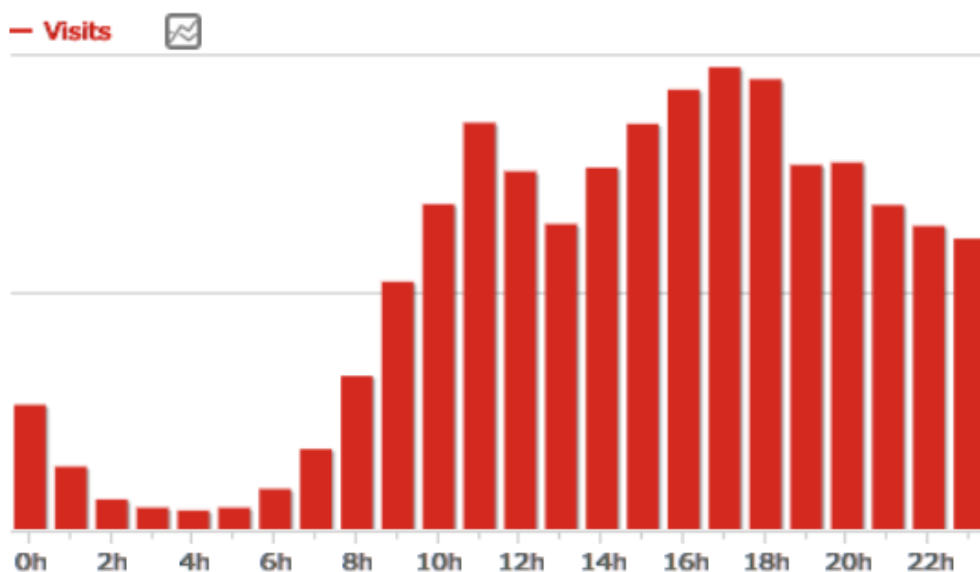
Actuellement, l'accès au site se fait essentiellement via des ordinateurs (92,5%). Parmi les autres modes de connexion, nous savons que 4,7% des visites sont réalisées à partir d'une tablette et 2,4% via un smartphone.

QUELLES SONT LEURS HABITUDES DE VISITE ?

Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (ressources intéressantes qui ont su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information recherchée, clics hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu).

Parmi toutes les visites effectuées sur Search, plus de trois sur quatre durent plus d'une minute (77%). Un peu plus de la moitié durent plus de 10 minutes (53,2%) et plus d'un tiers durent plus de 30 minutes (32,3%). En ce qui concerne le nombre de pages vues, plus de 80% des visites comptabilisent trois pages vues ou plus (80,1%), et plus de 60 % des visites comptent plus de 10 pages vues (63,1%).

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 10h et 23h.



search.arch.be - Fréquentation par heure (selon l'heure du serveur)⁷²

D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour Search, nous savons que 69% des visites se font en accès direct, c'est-à-dire que la personne a soit enregistré l'adresse du catalogue en ligne dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Un peu moins de 20% des visites (18,3%) sont le fruit d'une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduit au catalogue Search pourrait être très instructif. Malheureusement la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Nous avons ainsi plus de 21 000 mots-clés « not defined ». Les mots-clés disponibles ayant occasionné plus de 10 visites sont : « arch.be » (19 visites) ; « rijksarchief » (16) ; « zoeken in het rijksarchief » (15 visites) et « search.arch.be » (13 visites).

⁷² L'échantillon de données d'analyse pour Search est différent cette fois-ci : il concerne des données allant de janvier à décembre 2016 (raisons techniques).

Plus de 10% des visites (12,1%) correspondent à un clic sur un lien vers le catalogue Search effectué sur d'autres sites web/réseaux sociaux. Le site générant le plus de clics vers le catalogue Search est arch.be, qui génère 6,7% de toutes les visites. Les clics proviennent majoritairement des pages d'accueil et de celles consacrées aux généalogistes, dans leur différente version linguistique.

Enfin, la newsletter de l'institution est à l'origine de 0,6% des visites.

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ses forces et ses faiblesses.

En ce qui concerne les pages d'entrées les plus fréquentes, nous trouvons la page d'accueil sans précision de langue : search.arch.be. C'est sur cette page que commence 16,2% des visites. Parmi les autres pages générant beaucoup de visites, nous retrouvons en tête les pages de résultats du moteur de recherches par archives ainsi que les pages renvoyant aux registres paroissiaux⁷³ et celles renvoyant aux actes d'état civil⁷⁴ dans leur différente version linguistique.

⁷³ <http://search.arch.be/fr/tips/98-registresparoissiaux>

⁷⁴ <http://search.arch.be/fr/tips/101-etat-civil>

VARIA

Enfin, vu qu'en marge des fonctions de recherche dans le catalogue Search propose également plusieurs pages de contenu, nous avons mis en place une série de « goals »⁷⁵ afin d'évaluer leur succès⁷⁶.

| Goal | URLs correspondantes | Taux de conversion |
|--|---|--------------------|
| Contenus 1 : consultation d'archives numérisées | URLs débutant par <i>http://search.arch.be/./rechercher-des-archives/ /inventaris/</i> (+ version équivalente en NL, EN, DE) et contenant ensuite une mention de rabscan/eadid/... Exemple : <i>http://search.arch.be/fr/rechercher-des-archives/resultats/inventaris/rabscan/eadid/BE-A0541_006757_006696_DUT/inventarisnr/IRS000059461268/level/f ile/scan-index/158/foto/541_9999_999_696823_000_0_0158</i> | 54,7% |
| Contenus 2 : informations et liens ⁷⁷ | Toutes les URLs dont la structure reprend les caractères « tips », exemple : <i>http://search.arch.be/fr/tips/98-registresparoissiaux</i> | 35,1 |
| Bibliothèque (rubrique) | <i>http://search.arch.be/fr/bibliotheque</i> <i>http://search.arch.be/nl/bibliotheek</i> <i>http://search.arch.be/de/bibliotheek</i> <i>http://search.arch.be/en/bibliotheek</i> | 1,2% |
| Tutoriels 1 : PDF ⁷⁸ | <i>http://search.arch.be/images/stories/doc/handleiding_zoeken_person en.pdf</i> <i>http://search.arch.be/images/stories/doc/manuel_recherche_personn es.pdf</i> | 1,0% |
| Tutoriels 2 : « FAQ » | <i>http://search.arch.be/fr/home/faq</i> <i>http://search.arch.be/fr/rechercher-des-personnes/faq</i> + les versions équivalentes en NL/EN/DE | 1,0% |
| EAD-PDF (inventaires) | URLs débutant par <i>http://search.arch.be/ead/pdf/...</i> et se terminant systématiquement par ...ead.pdf Exemple : <i>http://search.arch.be/ead/pdf/BEA0541_006757_006696_DUT.ead.pdf</i> | 0,9% |

⁷⁵ Pour le détail de la méthode des goals, se référer à la première partie du rapport.

⁷⁶ Nous proposons ici une moyenne, basée sur les taux de conversion d'avril à septembre 2016.

⁷⁷ Après réflexion, nous n'avons pas repris les « pages d'accueil » des trois moteurs de recherche (telles que *http://search.arch.be/fr/rechercher-des-archives*) qui sont avant tout un espace de passage pour accéder aux fonctionnalités de recherche.

⁷⁸ Malheureusement, il n'était techniquement pas possible d'également prendre en compte les autres aides et guides hébergés sur *extranet.arch.be* :



Aperçu de Search, catalogue (en ligne) des Archives de l'État
(<http://search.arch.be/>)

Extraction des données

Quand on effectue des recherches dans le catalogue en ligne Search, l'URL reflète l'action que nous sommes en train de réaliser. Ici par exemple, nous retrouvons dans l'URL la trace du mot-clé qui a été encodé (« bruxelles ») dans le moteur de recherche principal (« rechercher des archives »), ainsi que l'intervalle de temps demandé (à partir de l'année 1800, jusqu'à 1900) :

search.arch.be/fr/rechercher-des-archives/resultats/index/index/zoekterm/bruxelles/beginjaar/1800/eindjaar/1900/lang/fr

Ces URLs, générées automatiquement pendant la visite des utilisateurs sur Search, représentent des données précieuses pour explorer la façon dont ceux-ci recherchent des contenus.

Les URLs sont stockées parmi d'autres informations dans une base de données. Ces informations sont enregistrées automatiquement par le serveur web à chaque nouvelle action sur le site. En partant de la documentation fournie par iMinds, les données jugées pertinentes ont pu être extraites de la base de données Piwik via une requête MySQL.

Dans le cadre du projet MADDLAIN, l'idée était d'explorer en priorité le type de requêtes effectuées par les visiteurs et la façon dont les filtres sont utilisées. Dans un premier temps, nous avons effectué une requête SQL qui visait à n'extraire que les URLs comportant des mots-clés encodés par les utilisateurs. En raison de la durée limitée du projet, seule l'activité du moteur de recherche par archives a été analysée. Ainsi, les informations concernant les moteurs de recherche par personne ou par producteur ne sont pas présentées ici. Cela pourra faire l'objet de futures recherches.

La présence des mots-clés est signalée par une chaîne de caractères apparaissant systématiquement avant le(s) mot(s)-clé(s) (le mot-clé étant Bruxelles dans ce cas-ci) :

*search.arch.be/fr/rechercher-des-
archives/resultats/index/index/zoekterm/bruxelles/beginjaar/1800/eindjaar/1900/lang/fr*

Nous avons donc utilisé un filtre composé de « */index/index/zoekterm/* ». Cela a permis de réduire considérablement la quantité de données à traiter. Ainsi, pour un échantillon d'une période de 6 mois, le filtre nous a permis de passer d'un volumineux fichier de 13 Go (l'entièreté des données brutes correspondant à l'activité sur search.arch.be) à un fichier d'environ 30 Mo⁷⁹.

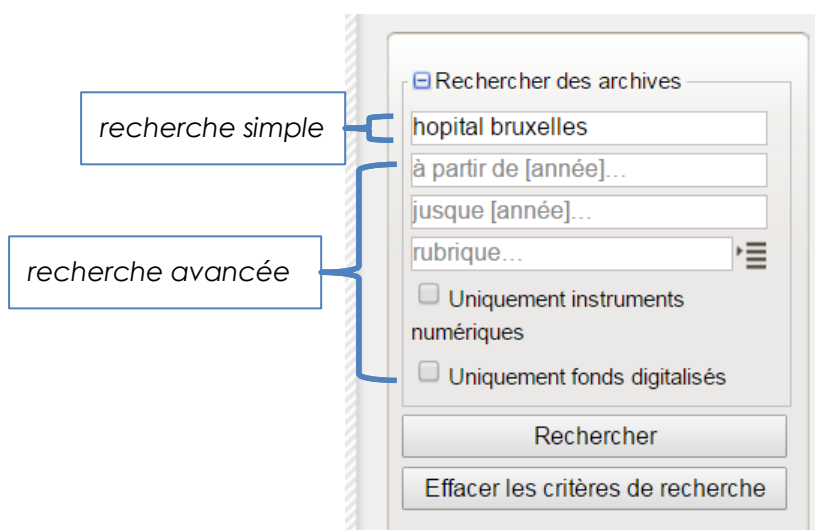
Recherche simple versus avancée

Il s'agit d'observer, lors de la recherche initiale, quel est le pourcentage de recherches où seul le champ « recherche libre » a été utilisé, par rapport à une « recherche avancée » (précisant une période dans le temps, une « rubrique » ou encore visant à faire apparaître uniquement les « instruments numériques » ou au les « fonds digitalisés »).

Nous avons utilisé un échantillon de données brutes correspondant à une période de 6 mois (janvier-juin 2016), dont le nombre total de requêtes s'élève à 127 763⁸⁰.

⁷⁹ Si ce système a permis d'obtenir de façon « artisanale » des données qui n'étaient jusque là pas disponibles, il possède néanmoins des limites. Ainsi, il est apparu que certaines expressions de recherche comportant des signes de ponctuation ou caractères spéciaux (exemples : « ; @ € () ¶ etc.) n'ont pas été prises en compte par le filtre utilisé.

⁸⁰ En présence d'une telle quantité de données, il arrive que certaines valeurs soient involontairement prises en compte. Ainsi, il est apparu au terme de l'analyse qu'un certain nombre d'URLs ne comportaient en réalité aucun



Type de recherche ⁸¹

| Type de recherche | Nombre ⁸² | En % |
|-----------------------|----------------------|--------|
| Recherche libre | 63 193 | 51,0 % |
| Recherche « avancée » | 60 686 | 49,0 % |

caractère dans l'espace où figure habituellement les mots encodés par l'utilisateur (il s'agit de 3 267 URLs sur un total de 127 763).

⁸¹ Ce type d'information peut être obtenu en utilisant des expressions régulières (par ex. `/(beginjaar|eindjaar|findaidstatus|dao)/` pour ne garder que les URLs constituant certaines caractéristiques.

⁸² Un certain nombre d'URLs contenues dans le corpus de départ (3 884) ne rentrait dans aucune des deux catégories lors de l'analyse effectuée via des expressions régulières. Il s'agit de « données sales », que l'on retrouve fréquemment dans de tels jeux de données, et dont il est difficile d'expliquer la cause.

Champs « dates »

| Année de début | Nombre de requêtes | En % |
|---|--------------------|------|
| champ laissé vide | 87 870 | 69 % |
| champ rempli (1268 « entrées » différentes) | 39 893 | 31 % |
| « Année de fin » | Nombre de requêtes | En % |
| champ laissé vide | 96 180 | 75 % |
| champ rempli (1268 « entrées » différentes) | 31 583 | 25 % |

TOP 10 « année de début »

| Rang | Valeur du champ | Nombre de requêtes |
|------|-----------------|--------------------|
| 1. | 1800 | 4 472 |
| 2. | 1700 | 2 761 |
| 3. | 1900 | 2 524 |
| 4. | 1600 | 1 808 |
| 5. | 1850 | 1 021 |
| 6. | 1500 | 808 |
| 7. | 1880 | 667 |
| 8. | 1750 | 547 |
| 9. | 1890 | 478 |
| 10. | 1914 | 454 |

TOP 10 « année de fin »

| Rang | Valeur du champ | Nombre de requêtes |
|------|-----------------|--------------------|
| 1. | 1900 | 4 281 |
| 2. | 1800 | 2 771 |
| 3. | 2000 | 1 651 |
| 4. | 1700 | 1 236 |
| 5. | 1850 | 769 |
| 6. | 1600 | 593 |
| 7. | 1950 | 591 |
| 8. | 1910 | 485 |
| 9. | 1920 | 453 |
| 10. | 1750 | 446 |

Par ailleurs, concernant les dates, l'analyse révèle que le champ est utilisé de manière très variée. En effet, vu qu'il n'y a pas de menus déroulants proposant des intervalles de temps préencodés, ni de contraintes formelles n'autorisant l'utilisateur à n'encoder que des chiffres, il s'avère que certaines personnes utilisent cette fonctionnalité de manière différente (cela reste cependant assez anecdotique en regard de l'activité de l'ensemble des utilisateurs). On retrouve ainsi toutes sortes d'expressions :

- **des années** (1592 ; 1622 ; ...)
- **des années écrites en toutes lettres** (mille huit centre trente quatre »)
- **des dates** (01.07.1760 ; 28 09 1918 ; 07-02-1814 ; ...)
- **des intervalles de temps** (1700-1800 ; 1947-1955 ; -1890 1910 ; 17001950 ; ...)
- **des expressions mélangeant des chiffres et des lettres** (« born 24 11 1879 » ; « Eeklo,23 maart 1873 » ; « geboorteakte 1822 » ; « 1830jusqu'au xxe S » ; « 1830-xxe s») « 1958, au Congo, province de Bandundu » ; « an II » ; « 1694 tot 1710 » ; « 1292 September 22 » ; ...)
- **des noms ou prénoms** (« joanna » ; « Lepas hortense eugenie » ; ...)
- **des localisations** (« brabant wallon » ; « turnhout » ; ...)
- **des (types de) contenus** (« etat civil » ; « procès » ; « mariages » ; « parochieregisters » ; « acte 311 » ; « matrice cadastrale 995 » ; ...)
- **divers** (« votre gestion » ; « colonial » ; ...)

Champs « rubrique »

| Rubrique | Nombre de requêtes | En % |
|-------------------|--------------------|------|
| champ laissé vide | 111 078 | 87 % |
| champ rempli | 16 685 | 13 % |

TOP 10 « rubrique »

| Rang | Rubrique | Nombre de requêtes |
|------|---|--------------------|
| 1. | Parochieregisters | 2 664 |
| 2. | Burgerlijke Stand (Gemeente en andere) | 1 790 |
| 3. | Gemeenten | 1 637 |
| 4. | Families | 978 |
| 5. | Particuliere personen | 793 |
| 6. | Parochiale instellingen, m.i.v. Heilige-Geestafel en kerkfabriek | 636 |
| 7. | Notarissen | 623 |
| 8. | Burgerlijke Stand (Rechtbank Eerste Aanleg) | 485 |
| 9. | Ambtenaren van de Burgerlijke Stand (+ sous-rubriques S1, S2, S3) | 336 |
| 10. | Stads- en gemeentebesturen | 326 |

« Uniquement instruments numériques » (case à cocher)

| Choix effectué par l'utilisateur | Nombre de requêtes | En % |
|---|--------------------|------|
| Case cochée (uniquement instruments numériques) | 15 468 | 12 % |
| Case non cochée | 112 295 | 88 % |

« Uniquement fonds digitalisés » (case à cocher)

| Choix effectué par l'utilisateur | Nombre de requêtes | En % |
|--|--------------------|------|
| Case cochée (uniquement fonds digitalisés) | 18 220 | 14 % |
| Case non cochée | 109 543 | 86 % |

« Langue de l'interface »

| Langue | Nombre de requêtes | En % |
|-------------|--------------------|--------|
| français | 70 564 | 55,2 % |
| néerlandais | 52 710 | 41,3 % |
| anglais | 3 316 | 2,6 % |
| allemand | 1 171 | 0,9 % |

« Archives par dépôt »

À l'aide de goals paramétrables directement dans l'interface graphique de Piwik⁸³, nous pouvons comparer de quels dépôts proviennent les archives recherchées⁸⁴.

Concrètement, nous utilisons les codes correspondants à chaque dépôt (exemple : Beveren = BE-A0512) qui sont présents dans les URLs d'affichage des résultats. De la même façon que pour les goals décrits en première partie de rapport, les goals (créés individuellement pour chaque dépôt) sont considérés comme « atteints » dès qu'une URL comportant certains caractères précis est visitée.

Pour des raisons pratiques, nous avons repris la liste des dépôts telle qu'elle apparaît sur le site dans la rubrique « filtres », bien qu'elle ne prenne pas en compte des changements qui sont survenus dans l'organisation physique des dépôts.

⁸³ Les taux de conversion présentés ici sont une moyenne issue des taux de conversion d'avril, mai, juin, juillet, août et septembre 2016.

⁸⁴ Bien entendu, ces chiffres doivent être considérés aux nombres d'archives mises en ligne par chaque dépôt.

| Dépôt | Taux de conversion |
|-------------------------------|---------------------------|
| Dépôt AGR | 2,3 % |
| Dépôt Antwerpen | 6,5 % |
| Dépôt Beveren | 5,2 % |
| Dépôt Brugge | 6,3 % |
| Dépôt Gent | 5,5 % |
| Dépôt Hasselt | 3,4 % |
| Dépôt Kortrijk | 4,3 % |
| Dépôt Ronse | 3,8 % |
| Dépôt Leuven | 6,4 % |
| Dépôt Arlon | 2,4 % |
| Dépôt Huy | 0,0 % |
| Dépôt Liège | 6,9 % |
| Dépôt Mons | 9,6 % |
| Dépôt Namur | 4,6 % |
| Dépôt Saint-Hubert | 1,2 % |
| Dépôt Tournai | 6,5 % |
| Dépôt Eupen | 0,8 % |
| Dépôt Anderlecht | 3,2 % |
| Dépôt LLN | 3,1 % |
| Dépôt AGR 2 – Joseph Cuvelier | 0,1 % |
| Dépôt Palais Royal | 0,1 % |

Trois moteurs de recherche : par archives / producteurs / personnes

À l'aide de goals paramétrables⁸⁵ directement dans l'interface graphique de Piwik⁸⁶, nous pouvons observer le succès relatif des trois moteurs de recherches présents sur *search.arch.be*.

| Moteur de recherche | Taux de conversion |
|---------------------|--------------------|
| Par archives | 50,2% |
| Par producteurs | 4,0% |
| Par personnes | 24,4% |

Afin d'obtenir des données représentatives de l'activité générale sur le catalogue et d'éviter des distorsions dues à l'activité d'utilisateurs particulièrement zélés qui auraient regardé plusieurs dizaines de pages de résultats (le mot-clé reste toujours présent dans l'URL qui « augmente » au cours du visionnage des résultats), nous avons décidé d'effectuer une réduction pour ne compter chaque requête distincte qu'une seule fois par visite. Cela a été possible grâce au `visit_ID` qui permet de rassembler les URLs par visite.

Voici les chiffres obtenus après nettoyage des données pour le moteur de recherche **par archives**, pour une période de six mois (janvier-juin 2016) :

| | |
|---------|------------------------------|
| Visites | Total requêtes ⁸⁷ |
| 42 046 | 67 426 |

⁸⁵ Les goals sont basés sur le contenu des URLs de la page d'affichage des résultats. Par exemple, pour le moteur de recherche par producteur, nous avons utilisé cette expression régulière : `http:\\\\search\\.arch\\.be\\(\\.\\)\\(producteurs-darchives|zoeken-naar-archiefvormers)\\(resultats|zoekresultaat).*`

⁸⁶ Les taux de conversion présentés ici sont une moyenne issue des taux de conversion de mai, juin, juillet, août et septembre 2016.

⁸⁷ Après réduction, en ne comptant chaque requête différente maximum une fois par visite.

D'autre part, les données disponibles ont également permis d'observer si l'utilisateur lambda effectue beaucoup de requêtes différentes au cours de sa visite sur Search. Il s'avère que 69,43% des visites (29 193 / 42 046) ont mené à la formulation d'une seule requête⁸⁸.

Moyenne de requêtes par visite : 1,6

Nombre maximum de requêtes différentes par visite : 8

Requêtes les plus fréquentes

Les requêtes encodées par les utilisateurs étant désormais accessibles, il est possible de savoir quels sont les termes qui reviennent le plus fréquemment.

Il est rapidement apparu que de légères nuances dans l'écriture d'un même mot ne permettraient pas de les comptabiliser comme une seule et même entité lors d'analyses automatisées (un être humain peut facilement associer « ACTE DE NAISSANCE » et « acte de naissance » mais pour une machine ce sont de deux résultats strictement différents). Ces nuances d'écriture peuvent être dues à des fautes de frappe, à l'usage de majuscules, de signes de ponctuation, etc. Pour régler cette question, il est possible de « normaliser » les chaînes de caractères de façon automatisée : on va ainsi tout mettre en minuscules, supprimer les signes de ponctuation, supprimer les espaces vides supplémentaires, etc.

De façon un peu plus expérimentale, un algorithme de « clustering » a également été utilisé. Ce dernier permet, en présence de grandes quantités de données, de rassembler des mots qui sont similaires sans être identiques. Il est donc possible d'associer des termes voisins tels que « collaboratie » ; « collaborateurs » ; « collobarateurs hainaut » ; ou encore « collaboration ». L'usage d'un tel algorithme peut donc aider à cerner de grands motifs récurrents en quelques clics, là où le calcul manuel de pourcentages au sein d'un corpus comptant plusieurs milliers de termes représenterait une charge de travail très lourde.

⁸⁸ Nous parlons de requêtes distinctes : il arrive que la personne ait effectué plusieurs recherches avec le même mot, dans ce cas, il n'est compté qu'une fois par visite.

Cette technique très puissante possède toutefois certaines limites. En effet, elle nécessite un contrôle manuel a posteriori et a parfois conduit à des résultats peu concluants lorsqu'il s'agissait de courts noms de lieux ou de personnes. Dans ces cas de figure, il serait sans doute plus pertinent de recourir aux techniques de « Named Entity Recognition » (NER).

Au vu du potentiel mais également des limites du clustering, nous avons décidé de présenter ici en vis-à-vis les résultats obtenus au terme de ces deux étapes (normalisation et clustering).

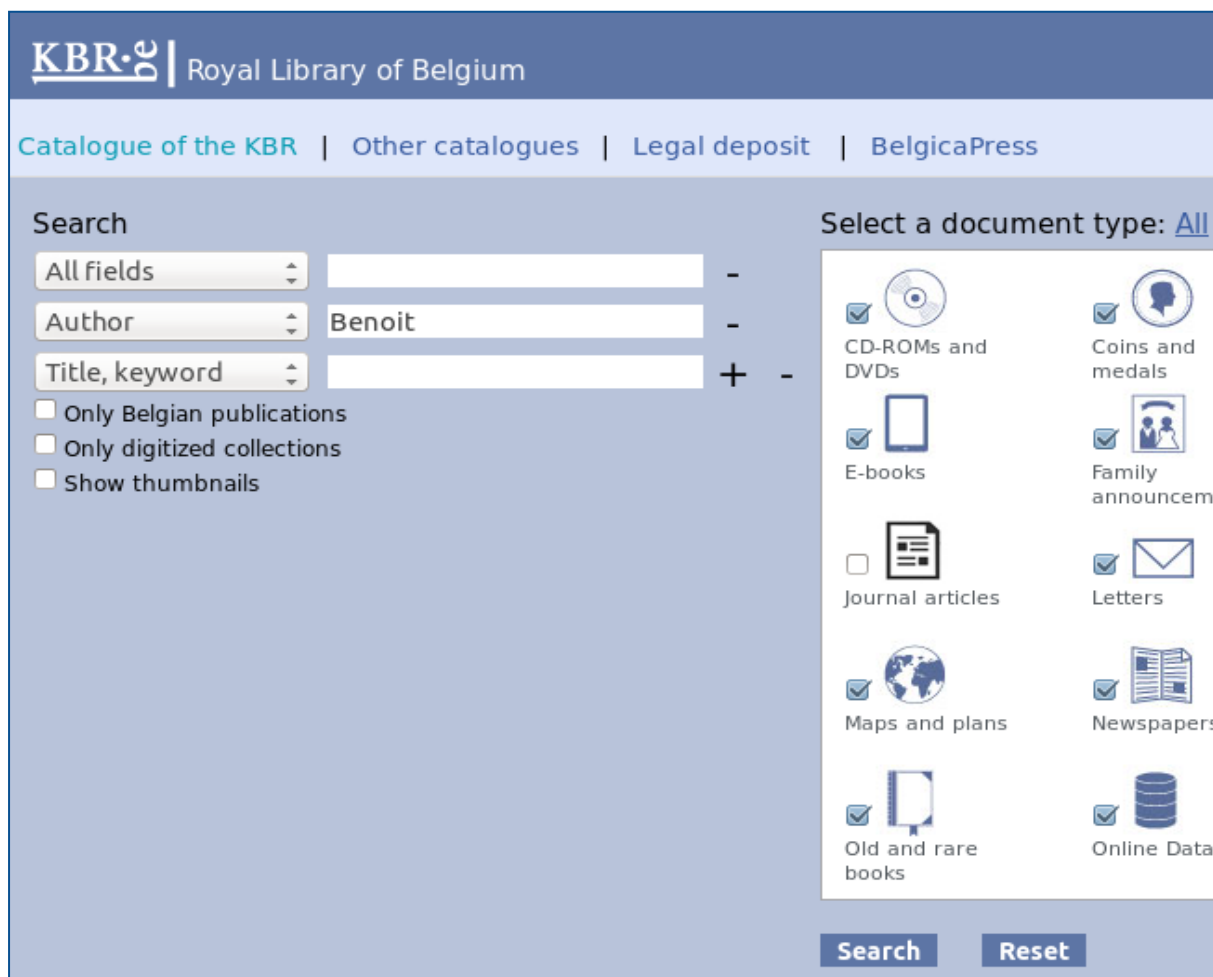
| | Après normalisation | Après clustering |
|---------------------------|---------------------|------------------|
| Total Requêtes distinctes | 37 399 | 21 724 |

20 requêtes apparaissant le plus souvent⁸⁹ :

| | Contenu de la requête après normalisation | Nombre occurrences | Contenu de la requête après clustering | Nombre occurrences |
|-----|--|---------------------------|--|---------------------------|
| 1. | etat civil | 394 | parochieregisters | 638 |
| 2. | parochieregisters | 370 | etat civil | 507 |
| 3. | registres paroissiaux | 285 | registres paroissiaux | 399 |
| 4. | rappports de paroisse | 206 | naissances | 259 |
| 5. | oorlogsverslagen | 182 | burgerlijke stand | 257 |
| 6. | digital | 179 | notaris | 236 |
| 7. | notaris | 176 | rappports de paroisse | 229 |
| 8. | naissance | 170 | acte de naissance | 213 |
| 9. | tournai | 163 | digital | 208 |
| 10. | archives du comite national de secours et d alimentation | 151 | mariage | 208 |
| 11. | liege | 140 | oorlogsverslagen | 204 |
| 12. | akten en contracten | 132 | tournai | 179 |
| 13. | namur | 129 | liege | 156 |
| 14. | mariage | 127 | archives du comite national de secours et d alimentation | 151 |
| 15. | commission d enquete sur la violation des regles du droit des gens | 124 | bruxelles | 134 |
| 16. | ellezelles | 120 | akten en contracten | 133 |
| 17. | oude universiteit leuven | 120 | namur | 132 |
| 18. | gent | 118 | ellezelles | 130 |
| 19. | kortrijk | 116 | antwerpen | 127 |
| 20. | contracten kortrijk | 114 | commission d enquete sur la violation des regles du droit des gens | 126 |

⁸⁹ En ce qui concerne la colonne « après clustering », il faut noter qu'un contrôle manuel a été réalisé au terme de l'opération pour exclure les résultats non pertinents.

2.3.2 Analyse spécifique : OPAC KBR



The screenshot shows the KBR (Royal Library of Belgium) OPAC search interface. At the top, the KBR logo and name are displayed. Below this, there are navigation links: "Catalogue of the KBR", "Other catalogues", "Legal deposit", and "BelgicaPress". The main search area is titled "Search" and contains three search fields: "All fields" (with a dropdown arrow), "Author" (containing "Benoit"), and "Title, keyword" (with a dropdown arrow). To the right of these fields are minus and plus signs. Below the search fields are three checkboxes: "Only Belgian publications", "Only digitized collections", and "Show thumbnails". To the right of the search area is a section titled "Select a document type: All" with a dropdown arrow. This section contains a grid of document type icons, each with a checkbox: CD-ROMs and DVDs (checked), E-books (checked), Journal articles (unchecked), Maps and plans (checked), Old and rare books (checked), Coins and medals (checked), Family announcements (checked), Letters (checked), Newspapers (checked), and Online Data (checked). At the bottom of the search area are two buttons: "Search" and "Reset".

COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 78 000 reprises.

Il est important de noter que parmi ces 78 000 visites, près de une visite sur cinq s'est terminée après que la visiteur n'a vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 18% pour le catalogue de la KBR.



Par ailleurs, il nous est possible de savoir que plus d'un visiteur sur deux est venu plus d'une fois sur le site (57,1%), 40% sont venus quatre fois ou plus et plus d'un visiteur sur dix est venu plus de 50 fois (10,4%).

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que plus de trois visites sur quatre (76,9%) sont effectuées depuis la Belgique, qui est suivie par la France (5,9%) et les Pays-Bas (4,3%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 40,7% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (18,6%), le Brabant flamand (6,4%), Anvers (4,4%), et la Flandre orientale (3,9%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (43,7%), suivis par le néerlandais (32,4%) ; viennent ensuite l'anglais (14,5%) et l'allemand (2,4%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (93,5%). Parmi les autres modes de connexion, nous savons que 3,8% des visites sont réalisées à partir d'une tablette et 2,4% via un smartphone.

QUELLES SONT LEURS HABITUDES DE VISITE ?

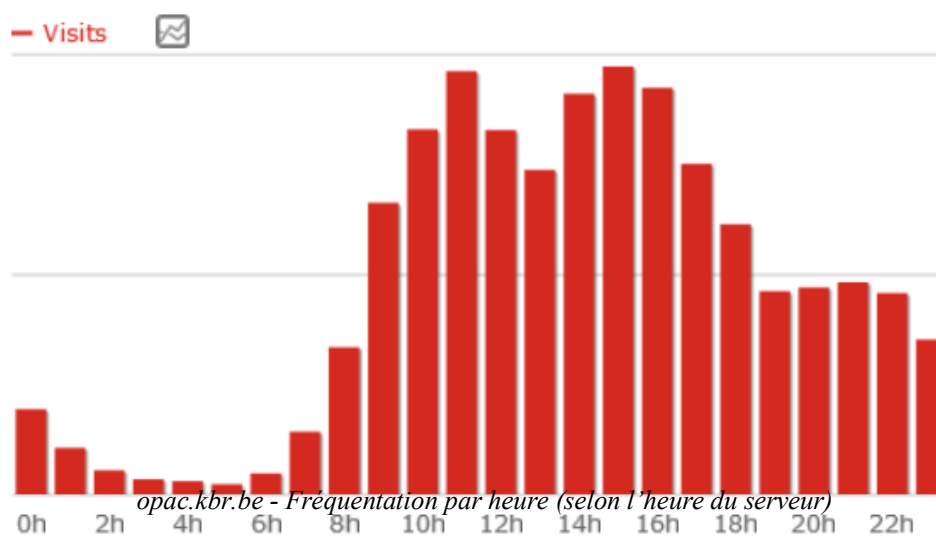
Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (ressources intéressantes qui ont su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information

recherchée, clics hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu).

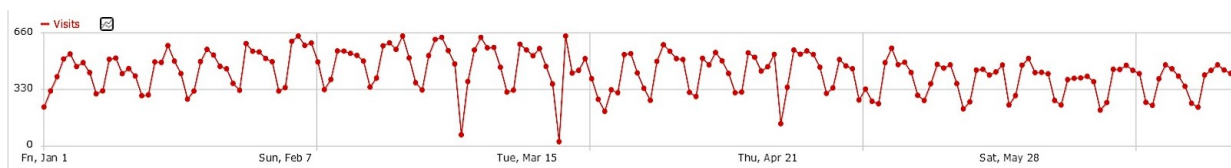
Parmi toutes les visites effectuées sur l'opac KBR, plus de 60% durent plus d'une minute. Un quart durent plus de 10 minutes (25,5%) et un dixième durent plus de 30 minutes (10,6%). En

ce qui concerne le nombre de pages vues, plus de trois visites sur quatre des comptabilisent trois pages vues ou plus (76,7%), et plus de 40 % des visites comptent plus de 10 pages vues (43,1%).

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 18h.



De plus nous pouvons également observer l'évolution des visites à travers le temps. Ainsi, nous voyons que les trois pics d'affluence les plus importants (640 visites) se situent autour du 9 février, du 25 février et du 21 mars 2016.



opac.kbr.be – Évolution des visites, jour par jour

Enfin, l'ajout de deux goals personnalisés pour l'opac KBR nous permet de savoir qu'environ un quart des visites comportent un passage sur BelgicaPress⁹⁰ (24,1% des visites entre le 1^{er} juin 2016 et le 1^{er} septembre 2016, soit 8 282 conversions sur un total de 34 292 visites) et que les pages d'aide⁹¹ sont visualisées au cours de 0,3% des visites seulement (107 conversions pour 34 292 visites).

D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour le catalogue de la KBR, nous savons que 35,5% des visites se font en accès direct, c'est-à-dire que la personne a soit enregistré l'adresse du catalogue en ligne dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Plus de 60% des visites (61,8%) correspondent à un clic sur un lien vers le catalogue effectué sur d'autres sites web/réseaux sociaux. Parmi les visites issues d'autres sites web, près de 95% proviennent du nom de domaine kbr.be (plus précisément de www.kbr.be (74,5%), qui est suivi de belgica.kbr.be (17,8%), pour ne citer que les sources principales). En outre, il est

intéressant de noter que sur toutes les visites effectuées sur le catalogue (78 510), plus d'un quart proviennent de la page d'accueil (en français ou néerlandais) du site kbr.be (respectivement 17,4% et 8,2%).

⁹⁰ Le goal reprend toutes les URLs commençant par opac.kbr.be/belgicapress.php...

⁹¹ Le PDF d'aide est disponible en anglais (exemple : *opac.kbr.be/help/opac_EN.pdf*), français et néerlandais.



Seuls 2,7% des visites sont le fruit d'une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduit au catalogue pourrait être très instructif. Malheureusement, la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Près de 98% des mots-clés sont ainsi « not defined ». Les mots-clés disponibles ayant occasionné le plus de visites sont : « bibliothèque royale bruxelles catalogue » (5 visites) ; « bibliothèque royale de belgique » (4 visites) ; « kbr catalogue en ligne » (4 visites) et « kbr catalogue » (3).

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ces forces et ces faiblesses.

En ce qui concerne les pages d'entrées les plus fréquentes, nous trouvons toute une variation d'URLs distinctes⁹² correspondant à la page d'accueil générale et à la page d'accueil de BelgicaPress, dans leurs différentes versions linguistiques⁹³. Si l'on rassemble ces différentes entrées, il apparaît que 35,6% des visites démarrent sur la page d'accueil en français, 17,3% sur la page en néerlandais et 7,3% sur une adresse de page sans langue spécifiée (exemple : opac.kbr.be). L'anglais est clairement en retrait (0,2%). En ce qui concerne BelgicaPress, plus de 17,3% des visites commencent sur les différentes versions linguistiques de cet onglet : 6,9% pour le néerlandais ; 6,0% pour le français ; 3,8% non précisé et 0,6% en anglais. Il faut noter que la page d'accueil néerlandophone de BelgicaPress génère plus de visites que la page francophone. Si la différence reste anecdotique (6,9% versus 6,0%), elle est cependant à contre-courant de la tendance générale qui a été observée dans le cadre de cette analyse (prépondérance du français sur le néerlandais).

⁹² Par exemple, Piwik classe « opac.kbr.be » et « opac.kbr.be/index.php » au sein de deux entrées différentes.

⁹³ Les autres pages d'entrée correspondent à des centaines d'URLs beaucoup plus longues et détaillées (par exemple le résultat d'une requête), sur laquelle n'ont démarré qu'un faible nombre de visites (ce qui n'est pas sans rappeler le phénomène de la « longue traîne »).



Aperçu du catalogue en ligne de la Bibliothèque royale de Belgique
(<http://opac.kbr.be/>)

Extraction des données

Quand on effectue des recherches dans l'opac KBR, l'URL reflète l'action que nous sommes en train de réaliser. Ici par exemple, nous retrouvons dans l'URL la trace du mot-clé qui a été encodé (« bruxelles ») :

```
http://opac.kbr.be/index.php?P0=QFICCAT&NR=&ALL=0&AA= Zoek  
&PAG=&BGSTR=&BBTEL=&HOSTNAME=&PW=&PINITP=0&P04100=&P07100=&P  
07000=10&P08100=&PDB=&BN=&HIDFILTER=&HIDSRTORD=&HIDSRT=&lang=FR  
&sd=0&select[]=P00000&searchfield[]=bruxelles&select[]=P01000&searchfield[]=&sele  
ct[]=P03000&searchfield[]=
```

Ces URLs, générées automatiquement pendant la visite des utilisateurs sur le catalogue, représentent des données précieuses pour explorer la façon dont ceux-ci recherchent des contenus.



Les URLs sont stockées parmi d'autres informations dans une base de données. Ces informations sont enregistrées automatiquement par le serveur web à chaque nouvelle action

sur le site. En partant de la documentation fournie par iMinds, les données jugées pertinentes ont pu être extraites de la base de données Piwik via une requête MySQL.

Dans le cadre du projet MADDLAIN, l'idée était d'explorer en priorité le type de requêtes effectuées par les visiteurs et la façon dont les filtres sont utilisés. Nous avons donc utilisé une requête MySQL visant à extraire toutes les URLs contenant les termes d'une requête effectuée dans le catalogue principal (les résultats ne concernent donc pas BelgicaPress).

La présence des mots-clés est signalée par une chaîne de caractères apparaissant systématiquement avant le(s) mot(s)-clé(s) (le mot-clé étant Bruxelles dans ce cas-ci) :

*[http://opac.kbr.be/index.php?P0=QFICCAT&NR=&ALL=0&AA= Zoek
&PAG=&BGSTR=&BBTEL=&HOSTNAME=&PW=&PINITP=0&P04100=&P07100=&P
07000=10&P08100=&PDB=&BN=&HIDFILTER=&HIDSRTORD=&HIDSRT=&lang=FR
&sd=0&select\[\]=P00000&searchfield\[\]=bruxelles&select\[\]=P01000&searchfield\[\]=&sele
ct\[\]=P03000&searchfield\[\]=](http://opac.kbr.be/index.php?P0=QFICCAT&NR=&ALL=0&AA=Zoek&PAG=&BGSTR=&BBTEL=&HOSTNAME=&PW=&PINITP=0&P04100=&P07100=&P07000=10&P08100=&PDB=&BN=&HIDFILTER=&HIDSRTORD=&HIDSRT=&lang=FR&sd=0&select[]=P00000&searchfield[]=bruxelles&select[]=P01000&searchfield[]=&select[]=P03000&searchfield[]=)*

Nous avons donc utilisé un filtre composé de « *searchfield* ». Cela a permis de réduire considérablement la quantité de données à traiter. En revanche, la structure des URLs est telle que nous n'avons pas pu extraire l'intégralité des données encodées au sein du moteur de recherche, qui permet de combiner plusieurs mots-clés (voir capture d'écran). Pour l'heure, seule une « ligne » de mots-clés a donc pu être extraite pour chaque visite⁹⁴.

⁹⁴ De plus, si ce système a permis d'obtenir de façon « artisanale » des données qui n'étaient jusque là pas disponibles, il possède néanmoins des limites. Ainsi, il est apparu au terme de l'analyse que certaines expressions de recherche comportant des signes de ponctuation ou caractères spéciaux (exemples : « ; @ € () ¶ etc.) n'ont pas été prises en compte par le filtre utilisé.

Rechercher dans

| | | |
|---------------------|---------------------------|---|
| Tous les champs | keyword 1 | - |
| Auteur | keyword keyword keyword 2 | - |
| Lieu de publication | keyword 3 | - |
| Titre, mot du titre | keyword 4 | - |

Uniquement publications belges
 Uniquement collections numérisées
 Afficher les miniatures

Afin d'obtenir des données représentatives de l'activité générale sur le catalogue et d'éviter des distorsions dues à l'activité d'utilisateurs particulièrement zélés, nous avons décidé d'effectuer une réduction pour ne compter chaque requête distincte qu'une seule fois par visite. Cela a été possible grâce au `visit_ID` qui permet de rassembler les URLs par visite. Voici les chiffres obtenus après nettoyage des données pour le moteur de recherche principal, pour une période de six mois (janvier-juin 2016) :

| Visites | Total requêtes ⁹⁵ |
|---------|------------------------------|
| 45 560 | 192 841 |

⁹⁵ Après réduction, en ne comptant chaque requête différente maximum une fois par visite.



D'autre part, les données disponibles ont également permis d'observer si l'utilisateur lambda effectue beaucoup de requêtes différentes au cours de sa visite sur Belgica. Il s'avère que 43,83% des visites (19 968 / 45 560) ont mené à la formulation d'une seule requête⁹⁶.

Moyenne de requêtes par visite : 4,3

Nombre maximum de requêtes différentes par visite : 594⁹⁷

Les requêtes encodées par les utilisateurs étant désormais accessibles, il est possible de savoir quels sont les termes qui reviennent le plus fréquemment.

Il est rapidement apparu que de légères nuances dans l'écriture d'un même mot ne permettaient pas de les comptabiliser comme une seule et même entité lors d'analyses automatisées (un être humain peut facilement associer « ACTE DE NAISSANCE » et « acte de naissance » mais pour une machine ce sont de deux résultats strictement différents). Ces nuances d'écriture peuvent être dues à des fautes de frappe, à l'usage de majuscules, de signes de ponctuation, etc. Pour régler cette question, il est possible de « normaliser » les chaînes de caractères de façon automatisée : on va ainsi tout mettre en minuscules, supprimer les signes de ponctuation, supprimer les espaces vides supplémentaires, etc.

De façon un peu plus expérimentale, un algorithme de « clustering » a également été utilisé. Ce dernier permet, en présence de grandes quantités de données, d'associer des mots similaires qui ne sont toutefois pas identiques. Ainsi, il est possible de rassembler des termes très proches tels que « ferrariskaart », « ferraris kaarten », « ferraris kaart » ou encore « feerarisklaarten ». Ils sont tous écrits de façon différente mais désignent le même item. L'usage d'un tel algorithme peut aider à cerner de grands motifs récurrents en quelques clics, là où le calcul manuel de pourcentages au sein d'un corpus comptant plusieurs milliers de termes représenterait une charge de travail très lourde.

⁹⁶ Nous parlons de requêtes distinctes : il arrive que la personne ait effectué plusieurs recherches avec le même mot, dans ce cas, il n'est compté qu'une fois par visite.

⁹⁷ Ce nombre fort élevé a fait l'objet d'un contrôle manuel : il s'agit de requêtes encodées au cours d'une visite qui a duré près de 7 heures (8h16-15h07).



Cette technique très puissante possède toutefois certaines limites. En effet, elle nécessite un contrôle manuel a posteriori et a conduit parfois à des résultats peu concluants lorsqu'il s'agissait de courts noms de lieux ou de personnes. Dans ces cas de figure, il serait sans doute plus pertinent de recourir aux techniques de « Named Entity Recognition » (NER).

Au vu du potentiel mais également des limites du clustering, nous avons décidé de présenter ici en vis-à-vis les résultats obtenus au terme de ces deux étapes (normalisation et clustering).

| | Après normalisation | Après clustering |
|---------------------------|---------------------|------------------|
| Total Requêtes distinctes | 139 552 | 81 476 |

20 requêtes apparaissant le plus souvent ⁹⁸ :

| | Contenu de la requête après normalisation | Nombre occurrences | Contenu de la requête après clustering | Nombre occurrences |
|-----|--|---------------------------|---|---------------------------|
| 1. | popp** | 1 137 | cap [nombre]* | > 2500 |
| 2. | bruxelles | 290 | bro 2016 [nombre] * | 1 518 |
| 3. | ferraris | 190 | popp** | 1 175 |
| 4. | hugo claus | 103 | a 2016 [nombres] * | 654 |
| 5. | le soir | 160 | bruxelles | 370 |
| 6. | ysaye | 99 | ferraris | 241 |
| 7. | belgique | 93 | le soir | 182 |
| 8. | antwerpen | 77 | belgique | 167 |
| 9. | claus hugo | 74 | vandermaelen | 134 |
| 10. | brussel | 73 | ded a 2007 [nombre]* | 132 |
| 11. | vandermaelen | 65 | brussel | 127 |
| 12. | popp philippe christian | 63 | ms iii [nombre]* | 127 |
| 13. | la libre belgique | 59 | s ii [nombre]* | 113 |
| 14. | congo | 58 | histoire | 112 |
| 15. | histoire | 58 | hugo claus | 112 |
| 16. | leuven | 55 | cl 11 222 a [nombre]* | 109 |
| 17. | liege | 55 | mus ms [nombre]* | 73 |
| 18. | bruegel | 53 | antwerpen | 103 |
| 19. | namur | 53 | ysaye | 103 |
| 20. | rubens | 52 | iii 99 [nombre]* | 97 |

*Cotes d'ouvrages ou documents

**Atlas cadastral

⁹⁸ En ce qui concerne la colonne « après clustering », il faut noter qu'un contrôle manuel a été réalisé au terme de l'opération pour exclure les résultats non pertinents.

Paramètres de recherche

Il s'agit d'observer, lors de la recherche initiale⁹⁹, de quelle façon sont utilisées les fonctionnalités avancées (préférences concernant le type de documents recherchés, activation de l'affichage de miniatures, ou encore clic pour qu'apparaissent uniquement les publications belges ou les collections numérisées).

The screenshot shows a search interface with the following elements:

- Rechercher dans:** Three dropdown menus for 'Titre, mot du titre', 'Auteur', and 'Titre, mot du titre'. The first dropdown contains the text 'keyword'.
- Filters:** Three checkboxes: 'Uniquement publications belges', 'Uniquement collections numérisées', and 'Afficher les miniatures' (checked).
- Choisir un type de document:** A grid of 15 document type options, each with a checkbox and an icon. The selected options are 'Documents iconographiques', 'Cartes et plans', and 'Imprimés anciens et précieux'. Other options include 'Bases de données en ligne', 'CD-ROM et DVD', 'E-books', 'Enregistrements sonores', 'Faire-part', 'Imprimés contemporains', 'Journaux', 'Lettres', 'Manuscrits', 'Monnaies et médailles', and 'Objets'.
- Buttons:** 'chercher' and 'effacer' buttons at the bottom.

Nous avons basé notre analyse sur un échantillon de données brutes correspondant à une période de 6 mois (janvier-juin 2016), dont le nombre total de requêtes s'élève à 709 316. À ce stade, il n'est pas encore possible d'énoncer davantage d'informations concernant les champs les plus utilisés pour ce qui est de « rechercher dans » (auteur, titre, ISBN, etc.).

⁹⁹ Nous avons utilisé un filtre (« Zoek&PAG=& ») pour ne garder que les URLs correspondant à la recherche initiale en excluant les pages de visualisation des résultats.

« Type de document »

| Type de document | Nombre de requêtes | En % |
|----------------------------|--------------------|-------|
| tous / aucun | 615 528 | 86,8% |
| sélection personnalisée | 92 906 | 13,1% |
| autres / erreurs | 882 | 0,1% |

Par ailleurs, si l'on fait abstraction du choix de « tous » / « aucun » type de document, voici les types les plus populaires (ordre décroissant) :

| Type de document (cat. non exclusives) | |
|--|-------|
| Journaux | 31636 |
| Imprimés contemporains | 21377 |
| Manuscrits | 21358 |
| Bases de données (en ligne) | 19131 |
| Documents iconographiques | 16981 |
| E-books | 16311 |
| Lettres | 16031 |
| Cartes et plans | 15601 |
| Faire-part | 14846 |
| CD-ROMs et DVDs | 14368 |
| Objets | 14003 |
| Musique imprimée et manuscrite | 13759 |
| Imprimés anciens et précieux | 13473 |
| Monnaies et médailles | 13452 |
| Enregistrements sonores | 8469 |



« Uniquement publications belges » (case à cocher)

| Choix effectué par l'utilisateur | Nombre de requêtes | En % |
|--|--------------------|--------------|
| Case cochée (uniquement publications belges) | 9 649 | 1 % (1,36) |
| Case non cochée | 699 667 | 99 % (98,64) |

« Uniquement collections numérisées » (case à cocher)

| Choix effectué par l'utilisateur | Nombre de requêtes | En % |
|---|--------------------|--------------|
| Case cochée (uniquement collections numérisées) | 20 306 | 3 % (2,86) |
| Case non cochée | 689 010 | 97 % (97,14) |

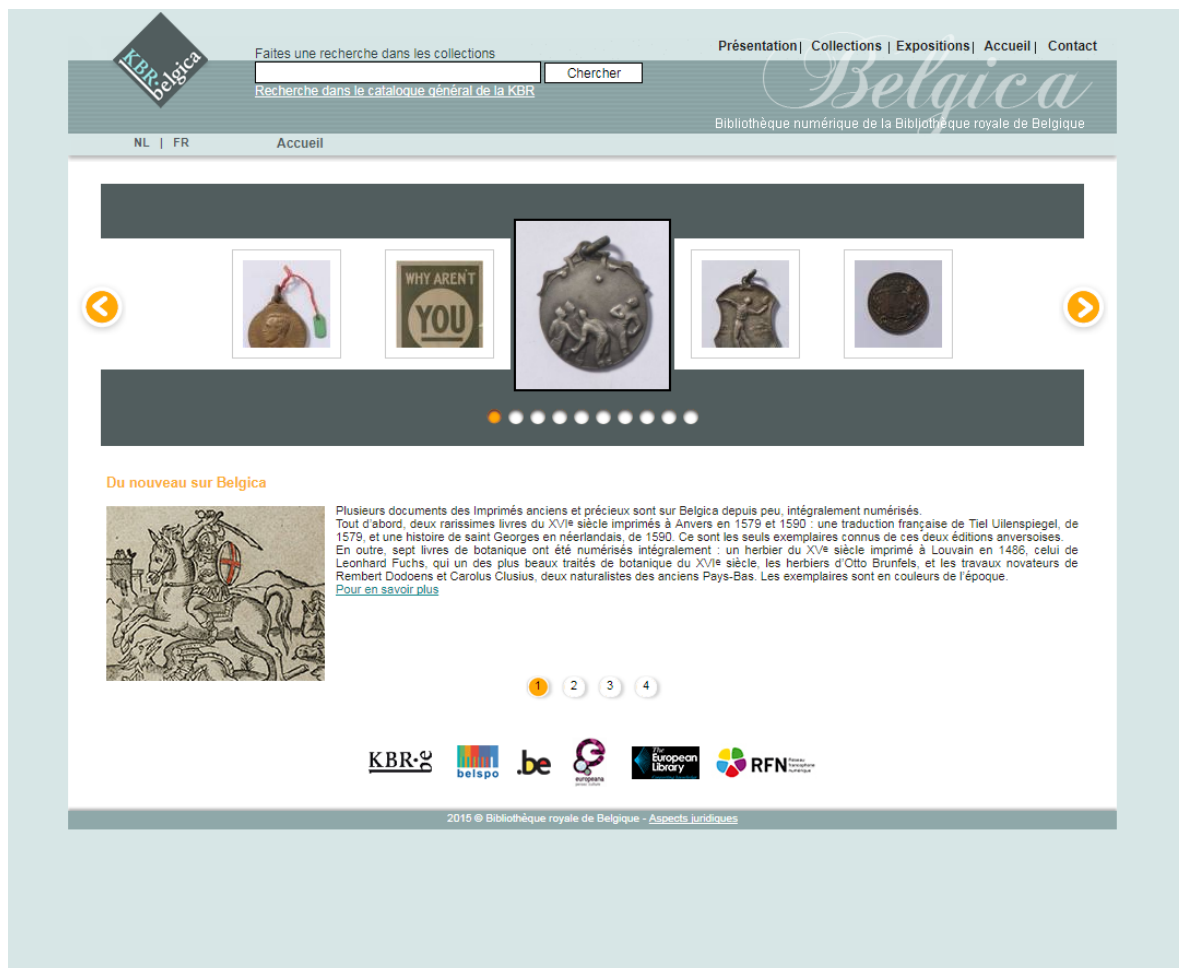
« Afficher les miniatures » (case à cocher)

| Choix effectué par l'utilisateur | Nombre de requêtes | En % |
|---------------------------------------|--------------------|--------------|
| Case cochée (afficher les miniatures) | 7 420 | 1 % (1,05) |
| Case non cochée | 701 896 | 99 % (98,95) |

« Langue de l'interface »

| Langue | Nombre de requêtes | En % |
|-------------|--------------------|------|
| français | 432 702 | 61 % |
| néerlandais | 269 825 | 38 % |
| anglais | 6 788 | 1 % |

2.3.3. Analyse spécifique : BELGICA



COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à près de 29 000 reprises¹⁰⁰.

Il est important de noter que parmi ces 29 000 visites, plus d'une visite sur trois s'est terminée après que le visiteur n'ait vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 36% pour Belgica.

¹⁰⁰ Pour être tout à fait précis, il faudrait ajouter à ce nombre les 721 visites effectuées sur opac.cegesoma.be durant la même période (si ce site et le catalogue Pallas sont étroitement liés, le site est toutefois considéré comme une entité distincte au sein de l'interface Piwik).



Par ailleurs, il nous est possible de savoir que seul un visiteur sur trois est venu plus d'une fois sur le site (32,1%), 16% sont venus quatre fois ou plus et moins de deux visiteurs sur 100 venus plus de 50 fois (1,4%).

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons qu'environ deux tiers des visites (64,7%) sont effectuées depuis la Belgique, qui est suivie par la France (9,6%) et les Pays-Bas (5,7%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 40,4% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (9,1%), Anvers (5,2%), la Flandre orientale (4,8%) et le Brabant flamand (4,8%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (38,9%), suivis de près par le néerlandais (36,3%). Viennent ensuite l'anglais (13,7%) et l'allemand (3,9%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (90,2%). Parmi les autres modes de connexion, nous savons que 5,9% des visites sont réalisées à partir d'une tablette et 3,3% via un smartphone.

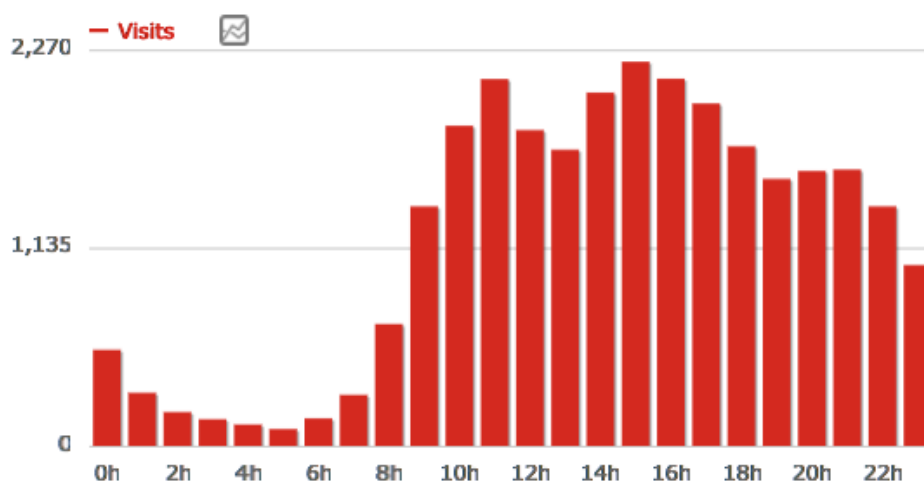
QUELLES SONT LEURS HABITUDES DE VISITE ?

Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visites), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (ressources intéressantes qui ont su captiver le

visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information recherchée, clics hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu).

Parmi toutes les visites effectuées sur Belgica, plus de la moitié durent plus de dix secondes (53,4%), mais moins d'un tiers durent plus d'une minute (31,8%). Un dixième des visites durent plus de 10 minutes (9,8%) et 1,8% durent plus de 30 minutes. En ce qui concerne le nombre de pages vues, un bon tiers des visites comptabilisent trois pages vues ou plus (37,4%), et seule une visite sur 20 compte plus de 10 pages vues (5,2%).

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 18h.



Belgica - Fréquentation par heure (selon l'heure du serveur)

De plus nous pouvons également observer l'évolution des visites à travers le temps. Ainsi, nous voyons trois pics d'affluence (comptabilisant plus de 225 visites) autour du 9, 22 et 29 février 2016.



Belgica – Évolution des visites, jour par jour



D'OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour la bibliothèque numérique Belgica, nous savons qu'un tiers des visites se font en accès direct (32,7%), c'est-à-dire que la personne a soit enregistré l'adresse du catalogue en ligne dans ses favoris, soit décidé d'en faire sa page d'accueil, soit tapé elle-même l'URL (ou le début de l'URL qui a été complété automatiquement) dans la barre d'adresse de son navigateur.

Près d'un quart des visites (23,0%) correspondent à un clic sur un lien vers le catalogue effectué sur d'autres sites web/réseaux sociaux. Parmi les visites issues d'autres sites web, un cinquième environ provient de opac.kbr.be, plus précisément de la page opac.kbr.be/catalog.php (18,1%) et de ses différentes versions linguistiques (1,4%). L'autre principale source de visites est www.kbr.be (19,3%) et notamment la page de la rubrique « collections » qui est consacrée aux catalogues.

Près de la moitié des visites (44,3%) sont le fruit d'une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduits au catalogue pourrait être très instructif. Malheureusement, la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Plus de 12 000 mots-clés (soit 95%) s'avèrent ainsi être « not defined ». Les mots-clés disponibles ayant occasionné plus de 10 visites sont : « belgica journaux » (32 visites) ; « ferrariskaart » (30 visites), « carte ferraris » (23 visites), « ferrariskaart online bekijken » (20 visites), « belgicapress » (16 visites), « ferraris » (16 visites), « ferrariskaart online » (15 visites), « kbr catalogue » (15 visites), « belgica » (14) et « belgica press » (11 visites).

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ses forces et ses faiblesses. Tout d'abord, à un niveau global, nous pouvons voir que 56,5% des visites se font sur une page en français, et 43,4% sur une page en néerlandais. Pour le français, les page générant le plus d'entrées sont la page d'accueil (3463 entrées), la page affichant la carte Ferraris (2411 entrées), la page ayant pour objet les journaux et plus précisément la numérisation de la presse belge (1942) et enfin la page consacrée à la Bibliotheca Carolina (974).

Pour le néerlandais, les pages générant le plus d'entrées sont la page d'accueil (2181), la page décrivant la carte Ferraris (2050) ainsi que la page affichant la carte Ferraris (1523) et enfin la page ayant pour objet les journaux et plus précisément la numérisation de la presse belge (1969).

VARIA

Enfin, nous avons mis en place une série de « goals »¹⁰¹ afin d'évaluer les types de pages de remportant le plus de succès¹⁰² sur Belgica.

| Goal | URLs correspondantes | Taux de conversion |
|---|---|--------------------|
| Contact / informations pratiques | http://belgica.kbr.be/fr/pres/pres_fr.html http://belgica.kbr.be/nl/pres/pres_nl.html http://belgica.kbr.be/fr/contact_fr.php http://belgica.kbr.be/nl/contact_nl.php | 2,1% |
| Liens externes vers des documents numérisés | URLs commençant par http://images.kbr.be/multi... http://dgtl.kbr.be:1801/view/action/singleViewer.do?dvs= http://images.kbr.be/mapview/index.php?image=/deventer Exemples : http://images.kbr.be/multi/KBR_5365Viewer/imageViewer.html http://dgtl.kbr.be:1801/view/action/singleViewer.do?dvs=1454328006237~95&locale=en_US&VIEWER_URL=/view/action/singleViewer.do?&DELIVERY_RULE_ID=10&frameId=1&usePid1=true&usePid2=true http://images.kbr.be/mapview/index.php?image=/deventer/1588248.imgf | 13% |

¹⁰¹ Pour le détail de la méthode des goals, se référer à la première partie du rapport.

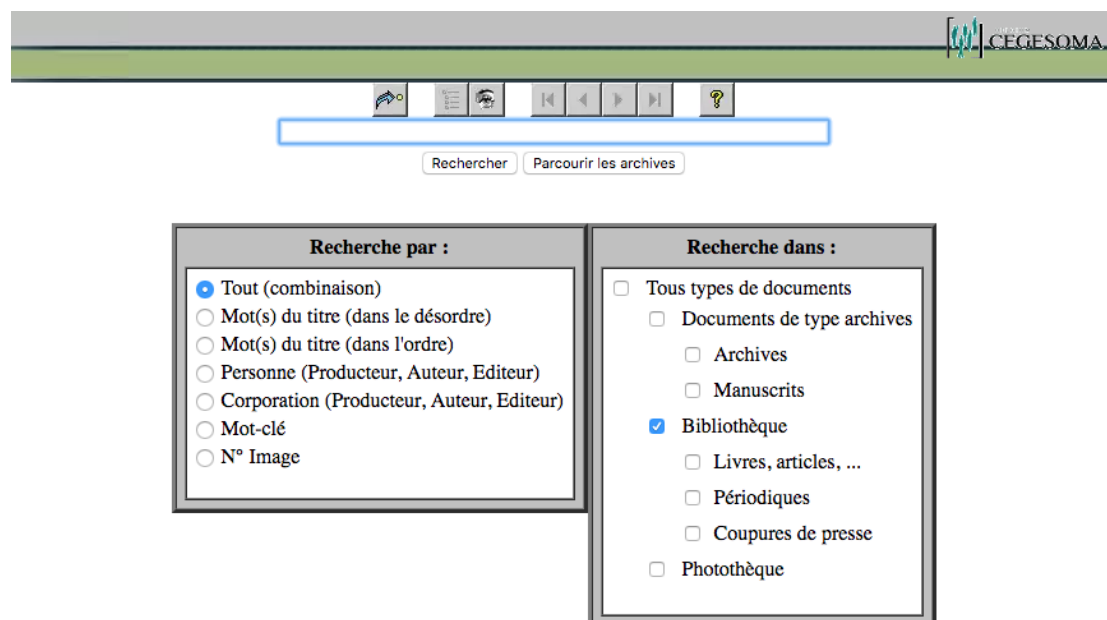
¹⁰² Nous proposons ici une moyenne, basée sur les taux de conversion de mars à août 2016 (01.03-01.09.2016).

| | | |
|---|---|--------------|
| <p>Informations sur les collections</p> | <p>http://belgica.kbr.be/fr/pres/pres_fr.html http://belgica.kbr.be/nl/pres/pres_nl.html http://belgica.kbr.be/fr/coll/jour/jour_fr.html http://belgica.kbr.be/nl/coll/jour/jour_nl.html http://belgica.kbr.be/fr/coll/jour/jourNumerisation_fr.html http://belgica.kbr.be/nl/coll/jour/gedigitaliseerdeKranten_n.html http://belgica.kbr.be/fr/coll/ouvRef/ouvRefHis_fr.html http://belgica.kbr.be/nl/coll/ouvRef/ouvRefHis_nl.html http://belgica.kbr.be/fr/coll/ouvRef/ouvRefCatal_fr.html http://belgica.kbr.be/nl/coll/ouvRef/ouvRefCatal_nl.html http://belgica.kbr.be/fr/coll/ouvRef/ouvRefExp_fr.html http://belgica.kbr.be/nl/coll/ouvRef/ouvRefExp_nl.html http://belgica.kbr.be/fr/coll/jour/presseArtistiqueLitteraire_fr.html http://belgica.kbr.be/nl/coll/jour/artistLitPers_nl.html http://belgica.kbr.be/fr/coll/jour/gazetteLeyde_fr.html http://belgica.kbr.be/nl/coll/jour/gazetteLeyde_nl.html.html</p> | <p>4,2%</p> |
| <p>Valorisation (clic via l'onglet Expositions)</p> | <p>http://vieuxtemps.kbr.be/fr/index.html http://ysaye.kbr.be/ http://www.kbr.be/fr/mozart http://www.kbr.be/fr/actualites/expos/virtuelles/tresors_inconables/index_fr.html http://bartok.kbr.be/expo_virtuelle/fonds_dille.html + liens équivalents en NL</p> | <p>0,5%</p> |
| <p>Lien vers le catalogue de la KBR</p> | <p>URLs commençant par http://opac.kbr.be/... ou http://opteron1.kbr.be/... ou http://dgtl.kbr.be:8881/R/A5XVGYPE.....</p> | <p>38,2%</p> |
| <p>Aide en ligne</p> | <p>http://belgica.kbr.be/fr/pres/pres_fr.html http://belgica.kbr.be/nl/pres/pres_nl.html http://belgica.kbr.be/fr/contact_fr.php http://belgica.kbr.be/nl/contact_nl.php http://belgica.kbr.be/fr/coll/jour/jourNumerisation_fr.html http://belgica.kbr.be/nl/coll/jour/gedigitaliseerdeKranten_n.html</p> | <p>4,7%</p> |
| <p>Carte Ferraris</p> | <p>http://belgica.kbr.be/fr/coll/cp/cpFerrarisCarte_fr.html http://belgica.kbr.be/fr/coll/cp/cpFerrarisCarte_nl.html</p> | <p>17,6%</p> |



| | | |
|--|--|-------------|
| <p>Métadonnées 1 (belgica.kbr.be)</p> | <p>http://belgica.kbr.be/fr/coll/ms/ms1312_fr.html http://belgica.kbr.be/fr/coll/ms/ms1791_94_fr.html http://belgica.kbr.be/fr/coll/ms/ms2921_22_fr.html http://belgica.kbr.be/fr/coll/ms/ms3641_fr.html http://belgica.kbr.be/fr/coll/ms/ms3701_15_fr.html http://belgica.kbr.be/fr/coll/ms/ms5169_fr.html http://belgica.kbr.be/fr/coll/ms/ms5253_fr.html http://belgica.kbr.be/fr/coll/ms/ms5325_27_fr.html http://belgica.kbr.be/fr/coll/ms/ms5374_75_fr.html http://belgica.kbr.be/fr/coll/ms/ms5413_22_fr.html http://belgica.kbr.be/fr/coll/ms/ms7984_fr.html http://belgica.kbr.be/fr/coll/ms/ms8216_18_fr.html http://belgica.kbr.be/fr/coll/ms/ms8550_51_fr.html http://belgica.kbr.be/fr/coll/ms/ms8780_93_fr.html http://belgica.kbr.be/fr/coll/ms/ms9311_19_fr.html http://belgica.kbr.be/fr/coll/ms/ms9403_fr.html http://belgica.kbr.be/fr/coll/ms/ms9581_95_fr.html http://belgica.kbr.be/fr/coll/ms/ms9625_26_fr.html http://belgica.kbr.be/fr/coll/ms/ms9641_fr.html http://belgica.kbr.be/fr/coll/ms/ms9845_48_fr.html http://belgica.kbr.be/fr/coll/ms/ms9850_52_fr.html http://belgica.kbr.be/fr/coll/ms/ms9987_91_fr.html http://belgica.kbr.be/fr/coll/ms/ms10054_56_fr.html http://belgica.kbr.be/fr/coll/ms/ms10066_77_fr.html http://belgica.kbr.be/fr/coll/ms/ms10127_44_fr.html http://belgica.kbr.be/fr/coll/ms/ms10796_800_fr.html http://belgica.kbr.be/fr/coll/ms/ms18723_fr.html http://belgica.kbr.be/fr/coll/ms/ms111763_fr.html http://belgica.kbr.be/fr/coll/ms/ms112206_fr.html http://belgica.kbr.be/fr/coll/ms/ms112572_fr.html http://belgica.kbr.be/fr/coll/ms/ms5365_fr.html http://belgica.kbr.be/fr/coll/ms/ms6409_fr.html http://belgica.kbr.be/fr/coll/ms/ms9125_fr.html http://belgica.kbr.be/fr/coll/ms/ms9505_06_fr.html http://belgica.kbr.be/fr/coll/ms/ms9507_fr.html http://belgica.kbr.be/fr/coll/ms/ms9961_62_fr.html http://belgica.kbr.be/fr/coll/ms/ms10319_fr.html http://belgica.kbr.be/fr/coll/ms/ms11200_fr.html http://belgica.kbr.be/fr/coll/ms/ms11201_2_fr.html</p> | <p>4,0%</p> |
| <p>Métadonnées 2 (au sein du catalogue en ligne de la KBR)</p> | <p>URLs débutant par <i>opteron1.kbr.be/cgi-bin</i> (ancien opac de la KBR) Exemple : <i>http://opteron1.kbr.be/cgi-bin/opac.cgi?P0=FKBR&P1=3_JAN&P9=&P5=20&P4=&P2=3&P3=R_BBH&P6=67_1450043</i></p> | <p>1,6%</p> |
| <p>Métadonnées 2 – PDF (au sein du catalogue en ligne de la KBR)</p> | <p>URLs débutant par <i>opteron1.kbr.be/</i> (ancien opac de la KBR) et se terminant par .pdf Exemple : <i>http://opteron1.kbr.be/manus/BELGICA/B034/esquisse_hist_orique.pdf</i></p> | <p>1,8%</p> |

2.3.4. Analyse spécifique : PALLAS



Aperçu de Pallas, catalogue (en ligne) du Cegesoma
 (<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htm>)

COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à plus de 9 000 reprises¹⁰³.

Il est important de noter que parmi ces 9 000 visites, une visite sur cinq s'est terminée après que la visiteur n'a vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 20% pour le catalogue Pallas.

¹⁰³ Pour être tout à fait précis, il faudrait ajouter à ce nombre les 721 visites effectuées sur opac.cegesoma.be durant la même période (si ce site et le catalogue Pallas sont étroitement liés, le site est toutefois considéré comme une entité distincte au sein de l'interface Piwik). En raison de limites d'ordre technique et pratique, ces dernières n'ont pas pu être prises en compte dans le cadre des analyses que nous détaillons ici.

Par ailleurs, il nous est possible de savoir que moins d'un visiteur sur deux est venu plus d'une fois sur le site (44,0%), près de 30% sont venus quatre fois ou plus (29,7%) et un peu plus d'un visiteur sur vingt est venu plus de 50 fois (6,4%).

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que plus de trois visites sur quatre (76,5%) sont effectuées depuis la Belgique, qui est suivie par la France (7,6%) et les Pays-Bas (4,1%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 41,0% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (9,3%), le Brabant flamand (9,3%), Anvers (8,6%), et la Flandre orientale (6,7%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (39,1%), suivis par le néerlandais (29,7%), l'anglais (23,2%) et l'allemand (4,1%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (92,8%). Parmi les autres modes de connexion, nous savons que 4,6% des visites sont réalisées à partir d'une tablette et 2,3% via un smartphone.

QUELLES SONT LEURS HABITUDES DE VISITE ?

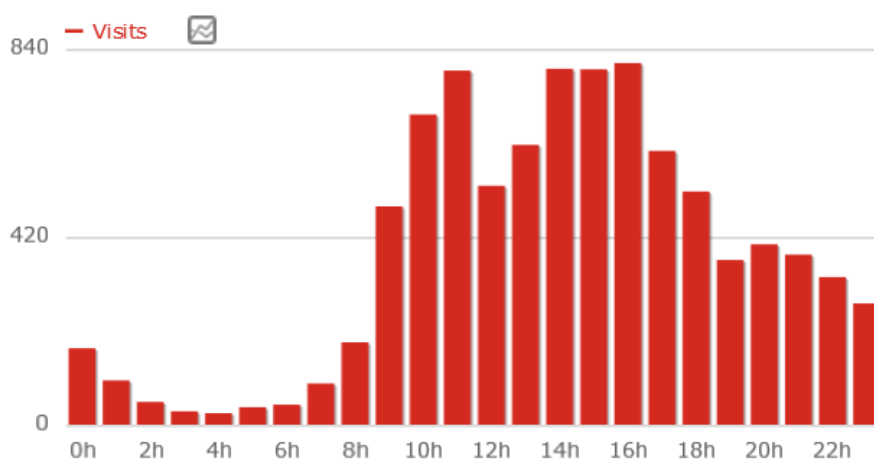
Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visite), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (ressources intéressantes qui ont su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information

recherchée, clics hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu).

Parmi toutes les visites effectuées sur Pallas, plus de la moitié durent plus d'une minute (53,1%). Un cinquième des visites durent plus de 10 minutes (19,8%) et 7,2% durent plus de

30 minutes. En ce qui concerne le nombre de pages vues, deux tiers des visites comptabilisent trois pages vues ou plus (67,8%), et près d'un tiers des visites comptent plus de 10 pages vues (31,1%).

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 9h et 18h.



Pallas (Opac Cegesoma) - Fréquentation par heure (selon l'heure du serveur)

De plus nous pouvons également observer l'évolution des visites à travers le temps. Ainsi, nous voyons deux grands pics d'affluence (comptabilisant plus de 90 visites) autour du lundi 7 mars et du mardi 15 mars 2016.



Pallas (Opac Cegesoma) – Évolution des visites, jour par jour

D’OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour le catalogue du Cegesoma, nous savons que 18,1% des visites se font en accès direct, c’est-à-dire que la personne a soit enregistré l’adresse du catalogue en ligne dans ses favoris, soit décidé d’en faire sa page d’accueil, soit tapé elle-même l’URL (ou le début de l’URL qui a été complété automatiquement) dans la barre d’adresse de son navigateur.

Près de 70% des visites (69,6%) correspondent à un clic sur un lien vers le catalogue effectué sur d’autres sites web/réseaux sociaux. Parmi les visites issues d’autres sites web, plus de 95% proviennent du nom de domaine cegesoma.be (plus précisément de www.cegesoma.be (92,9%) et de opac.cegesoma.be (2,4%)). En outre, il est intéressant de noter que sur toutes les visites effectuées sur le catalogue Pallas (9 151 visites), près d’un quart (24,1%) proviennent de la page « Catalogues » (en français, néerlandais ou anglais) du site cegesoma.be (respectivement 11,3%, 10,2% et 2,6%). Juste après nous trouvons la page intitulée « Archives » (sous-rubrique de la rubrique « Description des collections »), qui génère quant à elle 10,6% des visites sur Pallas (page en français : 5,8% ; néerlandais : 4,2% ; anglais : 0,6%).

Plus d’une visite sur dix (12,3%) est le fruit d’une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduits au catalogue pourrait être très instructif. Malheureusement, la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Près de 97% des mots-clés sont ainsi « not defined ». Les mots-clés disponibles ayant occasionné plus d’une visite sont : « cegesoma pallas » (5 visites) ; « pallas cegesoma » (5 visites) et « cegesoma » (4 visites).

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ces forces et ces faiblesses.

En ce qui concerne les pages d'entrées les plus fréquentes, nous trouvons sans surprise la page d'accueil de Pallas dans ses différentes versions linguistiques¹⁰⁴. Il apparaît que 30,7% des visites démarrent sur la page d'accueil en français¹⁰⁵, 24,7% sur la page en néerlandais¹⁰⁶, 7,8% sur la page en anglais¹⁰⁷ et 0,6% sur une page générale sans langue spécifiée¹⁰⁸, qui est par défaut en anglais.

Extraction des données

Quand on effectue des recherches sur Pallas, l'URL reste toujours la même :

<http://pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htm>

Pour pouvoir explorer la façon dont les utilisateurs recherchent des contenus, il a fallu ajouter des paramètres à ces URLs en les personnalisant. Ce travail a été pris en charge par les chercheurs d'iMinds. Ces derniers ont « augmenté » les URLs pour que l'on puisse savoir ce qu'un utilisateur recherche, les filtres qu'il utilise, etc. Exemple d'URL « augmentée » :

pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htmFormDatatext=Léon Degrelle La guerre en prison&action=search&Seop=5&in=_BA&

¹⁰⁴ Les autres pages d'entrée correspondent à des centaines d'URLs beaucoup plus longues et détaillées (par exemple le résultat d'une requête), sur laquelle n'ont démarré qu'un faible nombre de visites (ce qui n'est pas sans rappeler le phénomène de la « longue traîne »).

¹⁰⁵ pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=F&htdoc=general/opac.htm

¹⁰⁶ pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=N&htdoc=general/opac.htm

¹⁰⁷ pallas.cegesoma.be/pls/opac/plsp.getplsdoc?lan=E&htdoc=general/opac.htm

¹⁰⁸ pallas.cegesoma.be/pls/opac/plsp.getplsdoc?htdoc=general/opac.htm

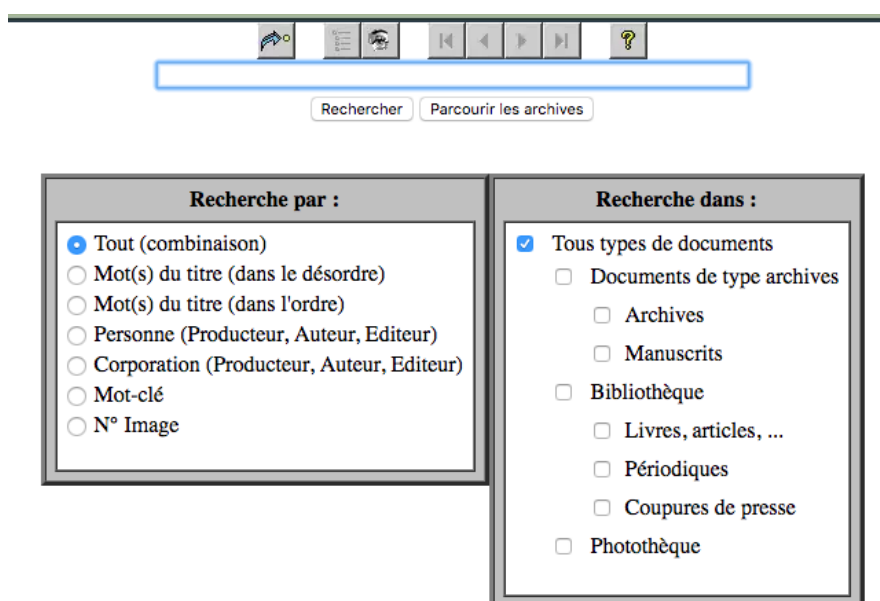
Ces URLs, générées automatiquement pendant la visite des utilisateurs sur Pallas, représentent des données précieuses pour explorer la façon dont ceux-ci recherchent des contenus.

Les URLs sont stockées parmi d'autres informations dans une base de données. Ces informations sont enregistrées automatiquement par le serveur web à chaque nouvelle action sur dans le catalogue. En partant de la documentation fournie par iMinds, les données jugées pertinentes ont pu être extraites de la base de données Piwik via une requête MySQL.

Dans le cadre du projet MADDLAIN, l'idée était d'explorer en priorité le type de requêtes effectuées par les visiteurs et la façon dont les filtres sont utilisées. Nous avons donc utilisé une requête MySQL visant à n'extraire que les URLs correspondant à la page d'accueil de Pallas, sur laquelle se trouve le formulaire de recherche dans le catalogue. Cela signifie que les informations concernant par exemple les pages de résultats ne sont pas reprises ici. Cela pourra faire l'objet de futures recherches.

Recherche simple versus avancée

Il s'agit d'observer, lors de la recherche initiale, quel est le pourcentage de « recherches simples » versus « recherches avancées ». Nous ne retrouvons pas les deux concepts tels quels au sein de Pallas, mais nous pouvons les reformuler en parlant de recherche « par défaut » (voir capture d'écran) et de recherche « personnalisée ».



The screenshot shows a search interface with a navigation bar at the top containing icons for search, home, and help. Below the navigation bar is a search input field with a blue border and two buttons: "Rechercher" and "Parcourir les archives".

Below the search field are two panels:

- Recherche par :**
 - Tout (combinaison)
 - Mot(s) du titre (dans le désordre)
 - Mot(s) du titre (dans l'ordre)
 - Personne (Producteur, Auteur, Editeur)
 - Corporation (Producteur, Auteur, Editeur)
 - Mot-clé
 - N° Image
- Recherche dans :**
 - Tous types de documents
 - Documents de type archives
 - Archives
 - Manuscrits
 - Bibliothèque
 - Livres, articles, ...
 - Périodiques
 - Coupures de presse
 - Photothèque

Nous avons utilisé un échantillon de données brutes correspondant à une période de 6 mois (janvier-juin 2016), dont le nombre total de requêtes s'élève à 24 865¹⁰⁹.

| Type de recherche | Nombre | En % |
|---------------------------|--------|-------|
| Par défaut ¹¹⁰ | 12 302 | 49,5% |
| Personnalisée | 12 563 | 50,5% |

¹⁰⁹ En présence d'une telle quantité de données, il arrive que certaines valeurs soient involontairement prises en compte. Ainsi, il est apparu au terme de l'analyse qu'un certain nombre d'URLs ne comportaient en réalité aucun caractère dans l'espace où figure habituellement les mots encodés par l'utilisateur (il s'agit de 3 267 URLs sur un total de 127 763).

¹¹⁰ Il s'agit de voir combien d'URLs comportent les éléments correspondant à Recherche par : All (combined) et Recherche dans « All document types ».

Rechercher versus parcourir les archives ?

| Action | Nombre | % |
|-------------------|--------|-----|
| « Search » action | 24 141 | 97% |
| « Browse | 724 | 3% |

« Recherche par »

| Champ(s) sélectionné(s) | Nombre | En % |
|---|--------|------|
| <i>All (combined)</i> | 17 810 | 72% |
| <i>Word(s) from title (random order)</i> | 630 | 3% |
| <i>Word(s) from title (as appearing)</i> | 286 | 1% |
| <i>Person (Creator, Author, Editor)</i> | 1 155 | 5% |
| <i>Corporate Body (Creator, Author, Editor)</i> | 27 | 0% |
| <i>Keyword</i> | 1 842 | 7% |
| <i>Nr Image</i> | 3 090 | 12% |

« Recherche dans »

| Champ(s) sélectionné(s) | Filtre présent dans <i>n</i> requêtes (catég. non exclusives) ¹¹¹ | En % | Nombre de visiteurs ayant utilisé ce filtre (catég. non exclusives) | En % |
|---------------------------------|--|---------|---|---------|
| All documents types | 17 596 | 71% | 2 047 | 85% |
| Archival documents | 316 | 1% | 115 | 5% |
| Archives | 562 | 2% | 155 | 6% |
| Manuscripts | 136 | 1% | 33 | 1% |
| Library | 879 | 4% | 119 | 5% |
| Books, articles, ... | 634 | 3% | 147 | 6% |
| Periodicals | 239 | 1% | 75 | 3% |
| Press Cuttings | 79 | 0% | 41 | 2% |
| Photo Library | 4516 | 18% | 635 | 26% |

Pour la page d'accueil du catalogue Pallas et pour une période de six mois (janvier-juin 2016), nous avons, après nettoyage des données :

| Visites | Requêtes ¹¹² |
|---------|-------------------------|
| 4 738 | 16 515 |

D'autre part, les données disponibles ont également permis d'observer si l'utilisateur lambda effectue beaucoup de requêtes différentes au cours de sa visite sur Pallas. Il s'avère que 42,76% (2 026/ 4 738) des visites ont mené à la formulation d'une seule requête.¹¹³

¹¹¹ Nous avons cherché à voir à combien de reprises le filtre est présent dans l'URL. Il s'agit de catégories non exclusives, vu que le filtre est parfois utilisé en combinaison avec d'autres filtres, voir vue détaillée en annexe.

¹¹² Après réduction, en ne comptant chaque requête différente maximum une fois par visite.

¹¹³ Nous parlons de requêtes distinctes : il arrive que la personne ait effectué plusieurs recherches avec le même mot, dans ce cas, il n'est compté qu'une fois.

Moyenne de requêtes par visite : 3,51

Nombre maximum de requêtes différentes par visite : 132

Les requêtes encodées par les utilisateurs étant désormais accessibles, il est possible de savoir quels sont les termes qui reviennent le plus fréquemment.

Il est rapidement apparu que de légères nuances dans l'écriture d'un même mot ne permettaient pas de les comptabiliser comme une seule et même entité lors d'analyses automatisées (un être humain peut facilement associer « ACTE DE NAISSANCE » et « acte de naissance » mais pour une machine il s'agit de deux résultats strictement différents). Ces nuances d'écriture peuvent être dues à des fautes de frappe, à l'usage de majuscules, de signes de ponctuation, etc. Pour régler cette question, il est possible de « normaliser » les chaînes de caractères de façon automatisée : on va ainsi tout mettre en minuscules, supprimer les signes de ponctuation, supprimer les espaces vides supplémentaires, etc.

De façon un peu plus expérimentale, un algorithme de « clustering » a également été utilisé. Ce dernier permet, en présence de grandes quantités de données, de rassembler des mots qui sont similaires sans être identiques. Il est donc possible d'associer des termes voisins tels que « collaboratie » ; « collaborateurs » ; « collobarateurs hainaut » ; ou encore « collaboration ». L'usage d'un tel algorithme permet de cerner de grands motifs récurrents en quelques clics, là où le calcul manuel de pourcentages au sein d'un corpus comptant plusieurs milliers de termes représenterait une charge de travail très lourde.

Cette technique très puissante possède toutefois certaines limites. En effet, elle nécessite un contrôle manuel a posteriori et a parfois conduit à des résultats peu concluants lorsqu'il s'agissait de courts noms de lieux ou de personnes. Dans ces cas de figure, il serait sans doute plus pertinent de recourir aux techniques de « Named Entity Recognition » (NER).

Au vu du potentiel mais également des limites du clustering, nous avons décidé de présenter ici en vis-à-vis les résultats obtenus au terme de ces deux étapes (normalisation et clustering).

| | | |
|---------------------------|---------------------|------------------|
| | Après normalisation | Après clustering |
| Total Requête différentes | 12 784 | 10 131 |

20 requêtes apparaissant le plus souvent¹¹⁴ :

| | Contenu de la requête après normalisation | Nombre d'occurrences | Contenu de la requête après clustering | Nombre d'occurrences |
|-----|--|-----------------------------|---|-----------------------------|
| 1. | sarajevo | 23 | collaboratie | 26 |
| 2. | degrelle | 22 | degrelle | 23 |
| 3. | rex | 21 | breendonk | 23 |
| 4. | congo | 18 | sarajevo | 23 |
| 5. | bruxelles | 17 | rex | 22 |
| 6. | algoet | 15 | stalag | 21 |
| 7. | repressie | 15 | antwerpen | 20 |
| 8. | antwerpen | 14 | prisonnier de guerre | 19 |
| 9. | gotovitch | 14 | repressie | 19 |
| 10. | hitler | 14 | bruxelles | 18 |
| 11. | armee secrete | 13 | congo | 18 |
| 12. | vnv | 13 | krijgsgevangene | 18 |
| 13. | breendonk | 11 | algoet | 17 |
| 14. | collaboration | 11 | armee secrete | 16 |
| 15. | dachau | 11 | bunker | 16 |
| 16. | legion wallonie | 11 | gotovitch | 15 |
| 17. | leuven | 11 | hitler | 15 |
| 18. | mons | 11 | oostende | 15 |
| 19. | resistance | 11 | de bruyne | 14 |
| 20. | robert verbelen | 11 | resistance | 14 |

¹¹⁴ En ce qui concerne la colonne « post clustering », il faut noter qu'un contrôle manuel a été réalisé au terme de l'opération pour exclure les résultats non pertinents.

2.3.5. Analyse spécifique : WARPRESS



The BELGIAN WAR PRESS

Rechercher...

English
Français
Nederlands

[Ce site web](#) [Consulter en ligne](#) [La presse de guerre](#)

The Belgian War Press vous emmène dans l'univers de centaines de journaux belges qui ont été rédigés, imprimés et diffusés dans le plus grand secret pendant les deux guerres mondiales. Jusqu'à ce jour, cette presse clandestine était dissimulée dans de multiples centres de documentation, dépôts d'archives et bibliothèques. Ces journaux ont été numérisés à l'initiative du CEGESOMA, grâce à un financement de Belspo. Vous pouvez de la sorte les lire et les parcourir aujourd'hui depuis votre domicile. C'est également valable pour la presse quotidienne censurée de la Première Guerre mondiale. The Belgian War Press vous offre aussi des informations sur le contexte de la presse de guerre censurée et clandestine. Et vous allez en apprendre plus sur le projet, ses partenaires et les principaux répertoires et ouvrages de référence.

COMBIEN DE VISITES ?

En l'espace de 6 mois (janvier à juin 2016), le site a été visité à un peu plus de 10 000 reprises¹¹⁵.

¹¹⁵ Pour être tout à fait précis, il faudrait ajouter à ce nombre les 721 visites effectuées sur opac.cegesoma.be durant la même période (si ce site et le catalogue Pallas sont étroitement liés, le site est toutefois considéré comme une entité distincte au sein de l'interface Piwik).

Il est important de noter que parmi ces 10 000 visites, plus d'une visite sur deux s'est terminée après que le visiteur n'ait vu qu'une seule page (la page via laquelle il est arrivé sur le site). Il s'agit du bounce rate, qui s'élève à 53% pour Belgica.

Par ailleurs, il nous est possible de savoir qu'environ un visiteur sur trois est venu plus d'une fois sur le site (29,0%), 16% sont venus quatre fois ou plus et un visiteur sur vingt sur 100 est venu plus de 50 fois (5,1%).

QUI SONT LES VISITEURS ?

En ce qui concerne le profil des visiteurs, il faut savoir que les statistiques proposées par Piwik sont généralement formulées en termes de visites plutôt que de visiteurs. Ainsi, contrairement à Google Analytics, Piwik ne nous dit par exemple rien sur l'âge ou le sexe des visiteurs. En revanche, nous possédons des informations sur le lieu et le mode de connexion au site.

Nous savons que plus de la moitié des visites (56,1%) sont effectuées depuis la Belgique, qui est suivie par la France (22,3%) et les Pays-Bas (4,0%). Parmi les visites provenant de Belgique, nous pouvons nous pencher sur les régions les plus représentées (outre les 38,2% dont on ne connaît pas l'emplacement exact) : Bruxelles-Capitale (9,1%), Anvers (5,2%), la Flandre orientale (4,8%) et le Brabant flamand (4,8%).

Si les données enregistrées par Piwik ne permettent pas de connaître de façon certaine la langue parlée par les visiteurs, des tendances sont toutefois perceptibles en identifiant la langue des navigateurs web : les visites sont majoritairement effectuées via des navigateurs configurés en français (42,8%), suivis par le néerlandais (33,7%), l'anglais (17,8%) et l'allemand (1,2%).

Actuellement, l'accès au site se fait essentiellement via des ordinateurs (86,5%). Parmi les autres modes de connexion, nous savons que 5,9% des visites sont réalisées à partir d'une tablette et 6,3% via un smartphone.

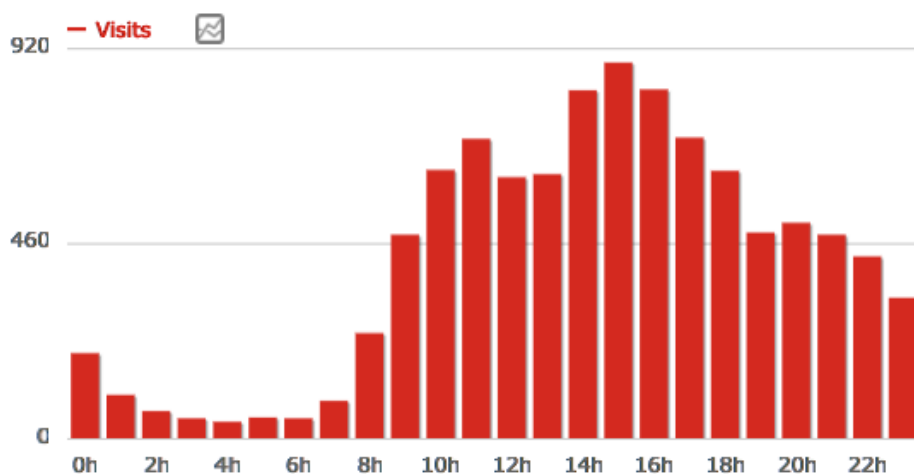
QUELLES SONT LEURS HABITUDES DE VISITE ?

Nous pouvons nous intéresser à la durée et à la profondeur de la visite (nombre de pages visionnées/visites), en gardant à l'esprit qu'une visite qui dure longtemps ou qu'un nombre de pages élevé peut être tant un signe de succès (ressources intéressantes qui ont su captiver le visiteur), qu'un symptôme de navigation laborieuse (difficulté à trouver l'information

recherchée, clics hasardeux ou obligation de parcourir plusieurs pages avant d'accéder à un certain contenu).

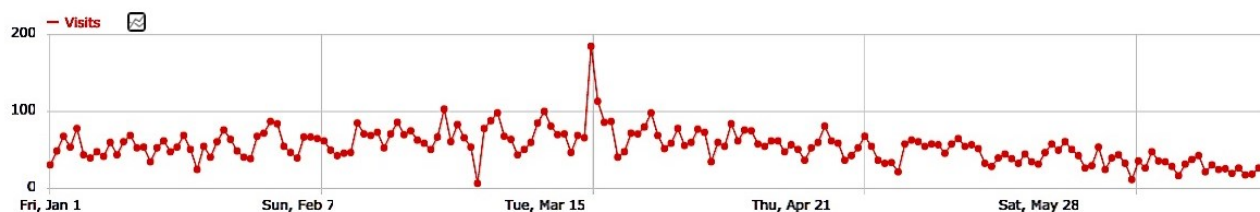
Parmi toutes les visites effectuées sur Wordpress, moins de la moitié durent plus de dix secondes (43,1%) et un tiers durent plus d'une minute (33,1%). Plus d'un dixième des visites durent plus de 10 minutes (13,7%) et 5,2% durent plus de 30 minutes. En ce qui concerne le nombre de pages vues, un tiers des visites comptabilisent trois pages vues ou plus (35,6%), et environ 15% des visites comptent plus de 10 pages vues (15,6%).

Les traces d'usage permettent également de repérer quelles sont les heures durant lesquelles il y a la plus grande affluence : entre 10h et 18h.



Wordpress - Fréquentation par heure (selon l'heure du serveur)

De plus nous pouvons également observer l'évolution des visites à travers le temps. Ainsi, nous voyons un important pic d'affluence (comptabilisant plus de 180 visites) autour du 22 mars 2016.



Warpress – Évolution des visites, jour par jour

D’OÙ VIENNENT-ILS ?

Il est possible de savoir comment un visiteur arrive sur le site. Ainsi, pour le site Warpress, nous savons que près d’un tiers des visites se font en accès direct (29,2%), c’est-à-dire que la personne a soit enregistré l’adresse du catalogue en ligne dans ses favoris, soit décidé d’en faire sa page d’accueil, soit tapé elle-même l’URL (ou le début de l’URL qui a été complété automatiquement) dans la barre d’adresse de son navigateur.

Plus d’un dixième des visites (12,9%) correspondent à un clic sur un lien vers le catalogue effectué sur d’autres sites web/réseaux sociaux. Parmi les visites issues d’autres sites web, plus d’un tiers proviennent de cegesoma.be : majoritairement des pages présentant le projet « The Belgian War Press » (les pages néerlandophones en tête), ainsi que de la sous-rubrique « Documents numérisés » (qui se trouve au sein de la rubrique « Trouver un document »).

Plus de la moitié des visites (57,9%) sont le fruit d’une requête lancée sur un moteur de recherche externe. Connaître les mots-clés des requêtes ayant conduit au catalogue pourrait être très instructif. Malheureusement, la plupart de ces visites sont associées au moteur de recherche de Google, qui garde cachés ces mots-clés depuis quelques années. Plus de 5 500 mots-clés (soit 95%) s’avèrent ainsi être « not defined ». Les mots-clés disponibles ayant occasionné le plus de visites sont : « gazet van brussel » (8 visites) ; « la presse clandestine » (5 visites), « presse clandestine seconde guerre mondiale » (5 visites), « cegesoma » (4 visites) et « la censure pendant la seconde guerre mondiale » (4 visites).

QUELLES SONT LES PAGES DU SITE AYANT UN HAUT TAUX D'ENTRÉE ?

Identifier ces pages peut fournir des indications en vue d'une amélioration du site, en identifiant ses forces et ses faiblesses. Tout d'abord, à un niveau global, nous pouvons voir que 44,1% des visites se font sur une page en français, et 36,6% sur une page en néerlandais. Pour le français, les pages générant le plus d'entrées sont celle sur la presse clandestine pendant la Seconde Guerre mondiale (1 293 entrées) et celle sur la presse censurée pendant la Seconde Guerre mondiale (781).

Pour le néerlandais, nous retrouvons parmi les pages générant le plus d'entrées la page sur la presse censurée durant la Seconde Guerre mondiale (314 entrée) ainsi que la page de la « Gazet van Brussel. Dagblad van het Vlaamsche Volk ».

VARIA

Enfin, nous avons mis en place une série de « goals »¹¹⁶ afin d'évaluer les types de pages¹¹⁷ remportant le plus de succès¹¹⁸ sur Wordpress.

| Goal | URLs correspondantes | Taux de conversion |
|--------------|--|--------------------|
| Contact | http://warpress.cegesoma.be/fr/contact http://warpress.cegesoma.be/en/contact http://warpress.cegesoma.be/nl/contact | 0,2% |
| Informations | http://warpress.cegesoma.be/fr/node/11 http://warpress.cegesoma.be/fr/node/12 http://warpress.cegesoma.be/fr/node/13 http://warpress.cegesoma.be/fr/node/17023 http://warpress.cegesoma.be/fr/node/6 http://warpress.cegesoma.be/fr/node/7 http://warpress.cegesoma.be/fr/node/8939 http://warpress.cegesoma.be/fr/node/8940 http://warpress.cegesoma.be/fr/node/96537 + pages équivalentes en EN et NL http://warpress.cegesoma.be/fr/newspaper-list?field_war_number_target_id=All&field_press_type_target_id=4 et toutes les autres pages similaires | 38,3% |

¹¹⁶ Pour le détail de la méthode des goals, se référer à la première partie du rapport.

¹¹⁷ Certains types de pages ne sont parfois pas *trackables* par Piwik, ce fut par exemple le cas des pages affichant des versions numérisées des journaux. Exemple :

<https://warpress.cegesoma.be/en/node/48099/download/kev9iga2Xgi9HAypjr2Q>

¹¹⁸ Nous proposons ici une moyenne, basée sur les taux de conversion du 1^{er} juin au 14 juillet 2016 (la brièveté de la période d'analyse est due à une interruption du tracking des visites sur Wordpress à partir du 15 juillet).

| | | |
|---------------------------------|---|-------|
| Valorisation | http://warpress.cegesoma.be/fr/node/11 http://warpress.cegesoma.be/fr/node/12 http://warpress.cegesoma.be/fr/node/13 http://warpress.cegesoma.be/fr/node/8 http://warpress.cegesoma.be/fr/node/8938 http://warpress.cegesoma.be/fr/node/8942 http://warpress.cegesoma.be/fr/node/8943 http://warpress.cegesoma.be/fr/node/8944 http://warpress.cegesoma.be/fr/node/8945 http://warpress.cegesoma.be/fr/node/8946 http://warpress.cegesoma.be/fr/node/8947 http://warpress.cegesoma.be/fr/node/8948 http://warpress.cegesoma.be/fr/node/8949 http://warpress.cegesoma.be/fr/node/8950 http://warpress.cegesoma.be/fr/node/8951 + pages équivalentes en EN et NL | 44,1% |
| Search (moteur de recherche) | http://warpress.cegesoma.be/fr/search/warpress http://warpress.cegesoma.be/fr/search/node | 14,6% |



Conclusion

La raison d'être de ce rapport était de rassembler tous les résultats obtenus dans le cadre du volet Web Analytics du projet MADDLAIN. Destiné à un usage interne, il présente les données brutes correspondant à la fréquentation des 12 sites web faisant partie du projet. Les données sont réparties au sein de deux parties distinctes : tout d'abord les sites web des institutions et ensuite les catalogues en ligne, sites d'accès aux collections numérisées et sites de valorisation. Dans la mesure du possible, une période d'analyse de 6 mois a été privilégiée, afin de de « lisser » les effets de contexte tels que la rentrée universitaire ou les vacances scolaires qui peuvent avoir un impact sur la fréquentation des sites web.

Allant de la répartition des utilisateurs par type de navigateur web utilisé au succès des filtres de recherche avancée, les résultats se veulent exhaustifs, notamment en vue de permettre au personnel des institutions de pouvoir réutiliser ces chiffres et pourcentages dans le cadre de la refonte d'un site web ou de leur travail de valorisation des collections.

Nous aimerions encore souligner que si ces données offrent de nouveaux éclairages sur les pratiques des utilisateurs et permettent de discerner de grandes tendances, elles restent toutefois baignées d'une certaine dose d'incertitude. En effet, de nombreuses limites se présentent dans ce type d'analyse parfois encore très expérimental. Ainsi, pour prendre un exemple très parlant, un même utilisateur se connectant à l'aide de trois appareils différents au cours de sa journée (dans les transports à l'aide de son smartphone, sur son lieu de travail à l'aide de son ordinateur professionnel et enfin à la maison sur sa tablette personnelle), sera malheureusement identifié par Piwik comme trois utilisateurs différents. À l'heure actuelle, il faut donc composer avec ces limites et considérer ces données davantage comme des indicateurs destinés à conduire à des prises de décision, plus que des données précises et parfaitement exactes.



Enfin, il est clair que l'exploration des données pourrait aller beaucoup plus loin encore. Pensons par exemple à la Bibliothèque nationale de France qui s'est alliée à l'école d'ingénieur ParisTech en vue de modéliser les comportements des utilisateurs à partir de l'analyse des logs de Gallica¹¹⁹. Ce sont des pistes qui pourront toujours être envisagées à l'avenir : bien que le projet MADDLAIN soit arrivé à son terme, les résultats obtenus, les données brutes et les outils de capture de données ont été archivés de manière à rester exploitables.

¹¹⁹ <https://blogrecherche.wp.imt.fr/2016/11/03/bnf-sociologie-big-data-gallica/>