

ANALYSE ET TENDANCES AVEC GOOGLE



LAURENT.KLING@epfl.ch, FACULTÉ STI

INTRODUCTION

Tel Janus, Internet présente deux faces:

- la mise à disposition du contenu
- l'analyse de l'utilisation.

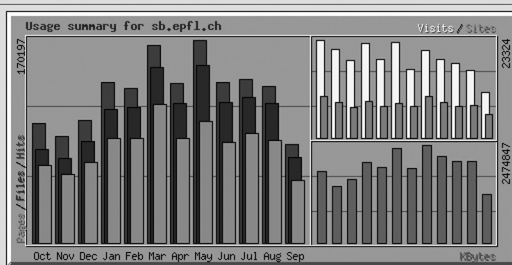
La création de pages Web est particulièrement bien documentée et accessible par de nombreux chemins, du code HTML écrit directement, à la conversion automatique de contenu.

L'analyse du résultat de la visite des pages Web est souvent délaissée faute d'outils réellement efficaces.

La première solution qui vient à l'esprit consiste à analyser le fichier de connexion du serveur Web. Ce travail peut être réalisé par un programme, Webalizer, qui a été décrit en mai 2004: http://ditwww.epfl.ch/SIC/SA/SPIP/Publications/article.php3?id_article=529. Par exemple, pour le site de la Faculté SB: <http://sb.epfl.ch/stats/>

Usage Statistics for sb.epfl.ch

Summary Period: Last 12 Months
Generated 21-Sep-2006 04:17 CEST



Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Sep 2006	3930	3411	2485	512	5670	1233632	10758	52192	71638	82545
Aug 2006	4221	3777	2781	519	7665	2061701	16114	86211	117087	130876
Jul 2006	4421	3942	2972	573	7551	2067394	17764	92162	122219	137059
Jun 2006	4492	3939	2800	626	8342	2176482	18784	84016	118173	134781
May 2006	5490	4788	3286	672	9889	2474847	20833	101879	148435	170197
Apr 2006	4467	3898	2924	546	7507	1890465	16409	87743	116960	134037
Mar 2006	5353	4736	3739	733	8286	2379571	22752	115929	146838	165970
Feb 2006	4615	4045	3121	664	7588	1900667	18619	87394	113269	129232
Jan 2006	4331	3603	2826	723	8628	2037127	22428	87632	111698	134276
Dec 2005	3313	2750	2187	595	7145	1598896	18452	67827	85273	102730
Nov 2005	2968	2374	1902	699	8503	1430687	20983	57088	71248	89065
Oct 2005	3240	2529	2090	752	9920	1806711	23324	64812	78414	100464
Totals					23058180	227220	984885	1301252	1511232	

Generated by Webalizer Version 2.01

Un décideur peut rester perplexé devant l'opacité des analyses fournies et leur manque de mise en relation.

Dans un songe, le cahier des charges d'un outil d'analyse devrait pouvoir offrir:

- le rapport entre nouveaux et anciens lecteurs et leurs nombres respectifs
- leur provenance
- leur langue
- l'analyse du parcours choisi

Et la variation de l'ensemble de ces analyses par rapport à une période précédente.

GOOGLE ANALYTICS

Ce rêve réalisé, c'est Google Analytics: <http://www.google.com/analytics/>. Le rachat de la société Urchin par Google a permis dans un premier temps de passer d'un abonnement mensuel de 495 \$ à 199 \$ pour cet outil. Maintenant, il est gratuit, avec uniquement une limite de 5'000'000 de pages vues par mois !

GOOGLE ANALYTICS DANS LE DÉTAIL

La seule opération nécessaire après son inscription est d'ajouter un fragment de code écrit en Javascript dans les pages dont on désire observer les fréquentations.

```
<script src="http://www.google-analytics.com/urchin.js" type="text/javascript">
</script>
<script type="text/javascript">
_uacct = "CouCou-1";
urchinTracker();
</script>
```

Ce code contient un identifiant, **CouCou-1**, lui permettant de distinguer un site d'un autre. Comme premières utilisations, j'ai ajouté ces codes dans trois sites:

- la Faculté STI.
- la section SGM.
- et le site concernant l'informatique pour la Faculté.

Après vérification de la présence de l'identifiant, on peut s'attendre à ce que les données soient analysées !

Une vue synthétique

Résumé des visites

Les quatre graphiques de ce rapport donnent un aperçu rapide des visites de votre site. Voici les éléments affichés :

- le nombre total de visites et de consultations de page enregistrées par votre site, le nombre moyen de consultations de page par visite (PVV) et le nombre de visites et de consultations de page par période. Les moyennes sont calculées en prenant l'ensemble des périodes sélectionnées, y compris les dates à venir lorsque cela est possible.
- le nombre de visiteurs accédant au site pour la première fois et de visiteurs déjà connus
- les villes d'origine des visiteurs les plus fréquents
- les sources principales.

Par défaut, Google Analytics offre un affichage de la dernière semaine écoulée. La sélection du jour, de la semaine et du mois permet de modifier l'analyse. Le même principe s'applique dans la comparaison entre deux périodes.

Cette vue offre:

- ▶ le nombre de visites et de pages consultées
- ▶ la proportion entre nouveaux et anciens visiteurs
- ▶ la répartition géographique des visites
- ▶ la provenance des visiteurs.

Une vue par objectif

Un objectif dans la vision Google correspond à la proportion d'utilisateurs qui atteignent la cible. Cette définition peut être affinée:

- ▶ une page Web
- ▶ un cheminement de page Web
- ▶ une expression régulière
- ▶ l'analyse d'un commerce électronique

Dans le cas de la Faculté STI, j'essaie de répondre à deux questions:

- quelle est la proportion de visiteurs qui utilise la version anglaise?
- quelle proportion de visiteurs utilise une représentation par nom ou par structure?

Quatre objectifs ont été choisis pour déterminer le trajet des internautes:

- les deux pages de garde, française et anglaise.
- les deux pages d'orientation dans les structures, nominatives et par acronymes.

Résumé des conversions				
stl.epfl.ch 11/09/2006 - 17/09/2006				
Conversion des objectifs	Visites	%±	Taux de con %±	
1. Visites	1 818	↓ -1%	100%	↓ 0%
2. G1: professeurs-alpha	121	↑ 1 629%	7%	↑ 1 649%
3. G2: professeurs-structure	57	↑ 1 800%	3%	↑ 1 823%
4. G3: home-english	671	↑ 1 216%	37%	↑ 1 232%
5. G4: home-francais	1 818	↑ 1 343%	100%	↑ 1 360%

Les résultats sont instructifs, 1/3 des internautes utilisent la page anglaise, mais après avoir été sur la page française.

Pour s'orienter dans les structures de recherche, la présentation nominative représente 60% des demandes, un résultat pour généraliser cette pratique dans les annuaires des structures de l'EPFL!

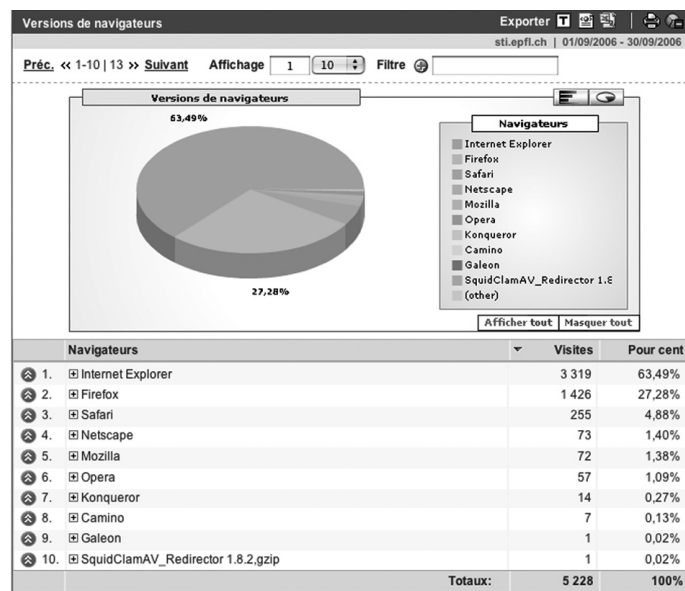
Avec la vue *marketing*, on possède une analyse croisée entre les sources, les objectifs et les mots-clés. On observe la part prépondérante de visites depuis les sites internes de l'EPFL, la faible importance des moteurs de recherche et un nombre non négligeable qui utilise directement l'adresse Web.

DES INDICATEURS TECHNIQUES

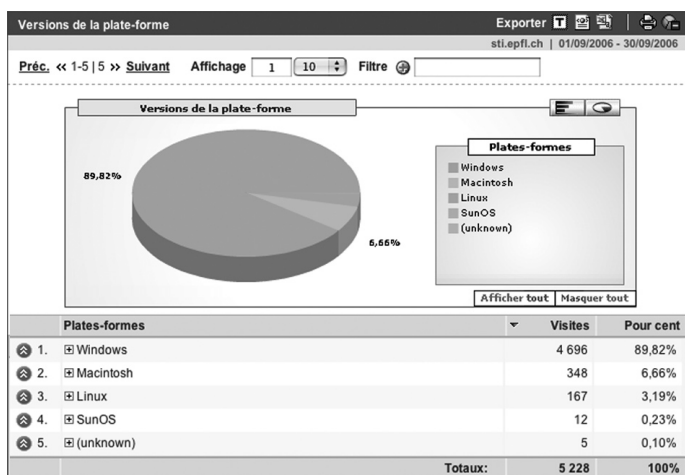
Les vues précédentes sont une vision synthétique de l'attractivité d'un site Web, il peut être utile d'observer de plus près les paramètres techniques sous-jacents aux choix technologiques. Pour minimiser les variations, une période d'un mois a été choisie dans les analyses suivantes.

Navigateurs utilisés

Internet Explorer représente la majorité des navigateurs, mais plus du tiers des internautes utilisent une autre plate-forme avec une prééminence de Firefox.



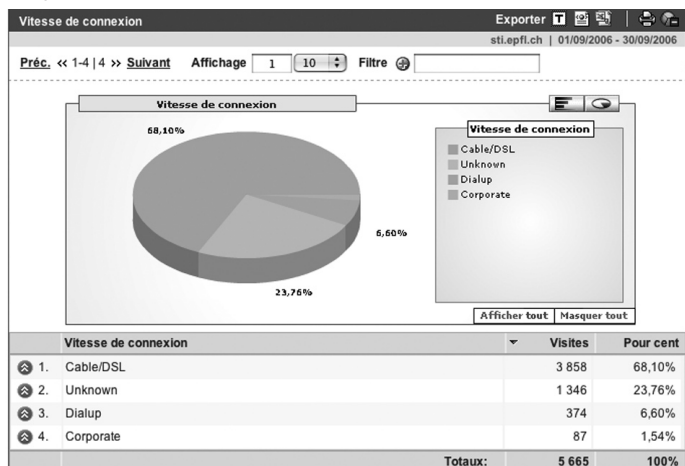
Plates-formes utilisées



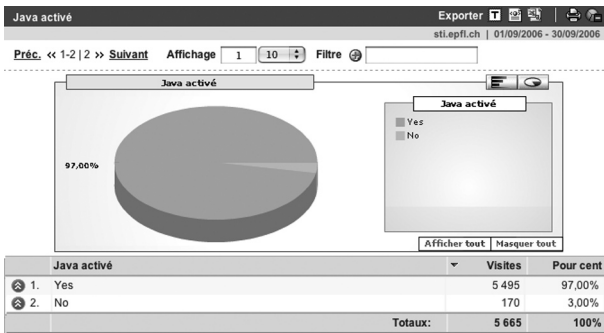
Les chiffres visualisent la répartition actuelle du marché. Le monde Macintosh pour 6.6% des internautes, Linux n'en représente que la moitié, 3.3%. Cette proportion faible montre que l'utilisation de Linux comme poste de travail bureautique n'est pas encore répandue.

Vitesse de connexion

Si 2/3 des internautes n'ont plus de problèmes de débit, l'optimisation pour la vitesse reste d'actualité pour 1/4 des usagers.

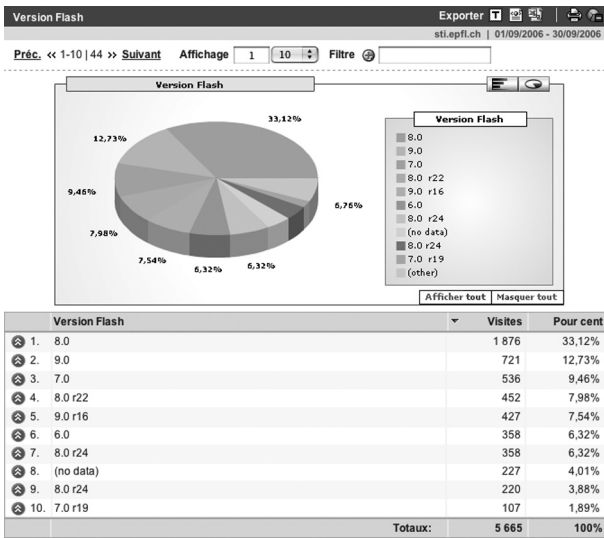


Java activé



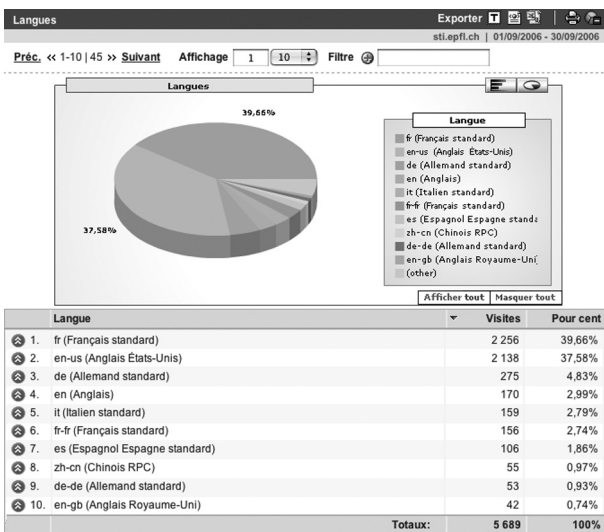
La démonstration n'est plus à faire, la quasi-totalité des internautes possède Java, soit 97%!

La version de Flash



Un casse-tête pour les créateurs de site avec la technologie Flash, si 1/3 des internautes utilisent la dernière version, le reste présente une grande biodiversité. Le problème de la compatibilité d'un site réalisé avec Flash se confirme!

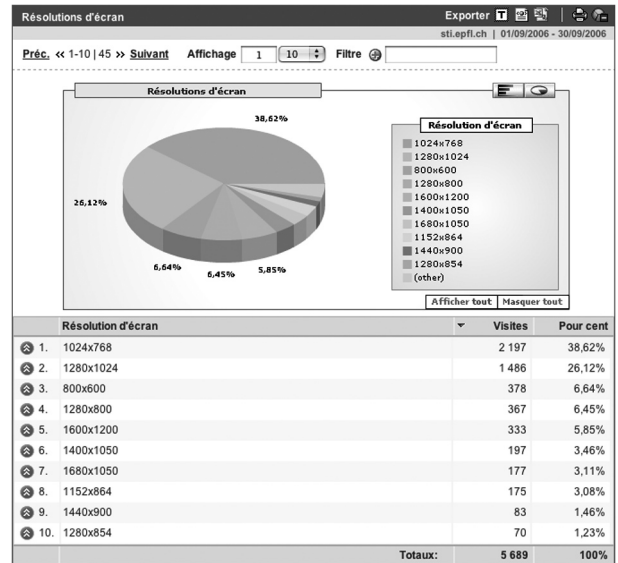
La langue utilisée



Un paradoxe est révélé par l'analyse de la langue fournie par le navigateur. Apparemment, une quasi-égalité règne entre l'anglais et le français. L'utilisation des objectifs montre

que seul un tiers des usagers se dirige vers la version anglaise. Ainsi, utiliser seulement l'indicateur de la langue du navigateur peut donner des résultats erronés.

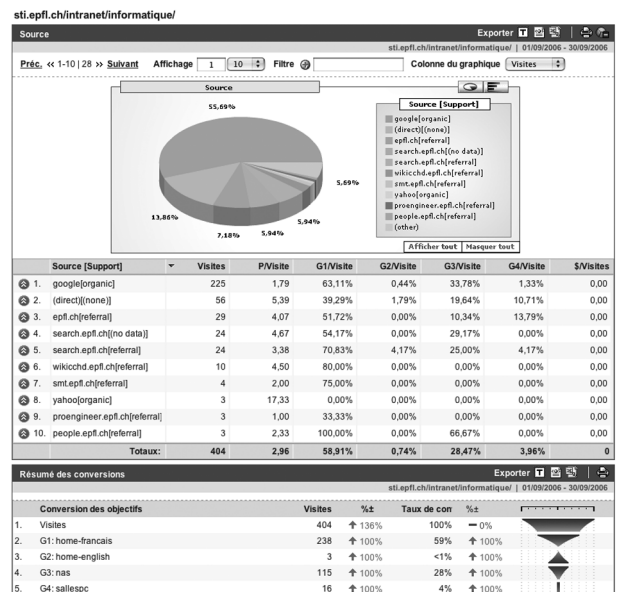
La résolution d'écran



La vision en 800 x 600 pixels semble appartenir au passé, elle ne représente plus que 7% des visites. Les résolutions de 1024 x 768 et de 1280 x 1024 doivent être le point de départ de la conception de sites Web. Si l'utilisation de résolution élevée reste minoritaire, elle est une part non négligeable du marché. Comme le pape de l'ergonomie Web Jakob Nielsen le préconise, une page à largeur fixe doit être bannie.

LE SITE D'AIDE POUR L'INFORMATIQUE DE LA FACULTÉ STI

Utilisation des moteurs de recherche



Le site institutionnel de la Faculté STI n'est que très peu visité par l'utilisation de moteur de recherche. La situation change complètement quand le site est orienté vers l'aide, par exemple l'aide pour les usagers pour l'informatique de la Faculté STI. En effet, plus de la moitié des visites vient depuis Google (contre 5% pour le site STI).

Comme l'ensemble des pages possède l'identifiant **Google Analytics**, il est intéressant de voir l'utilisation par objectif :

- G1 français
- G2 anglais

La proportion de visiteurs de la version anglaise est actuellement faible, mais la période observée est également brève!

- G3 NAS

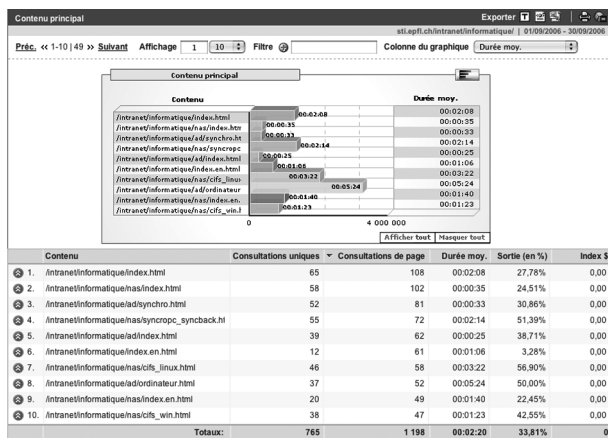
Une proportion importante d'internautes viennent lire les informations disponibles.

- G4 Salles PC

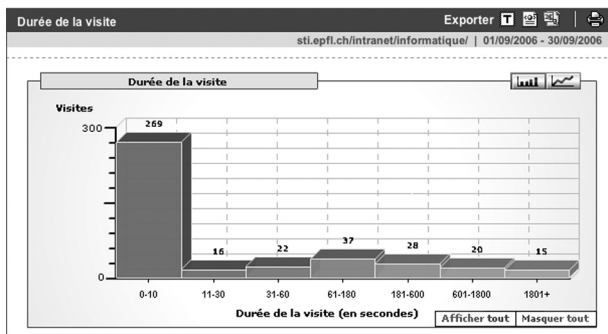
L'affluence des étudiants en octobre s'accompagnera probablement d'une augmentation similaire sur cette partie du site.

Durée de visite des pages

Un autre élément important est la durée de visite d'une page.



Pour un site d'explication, le temps passé à consulter une page est important, 2 min 20 en moyenne, la seule page *rapide* est l'index qui offre un rôle d'aiguillage.



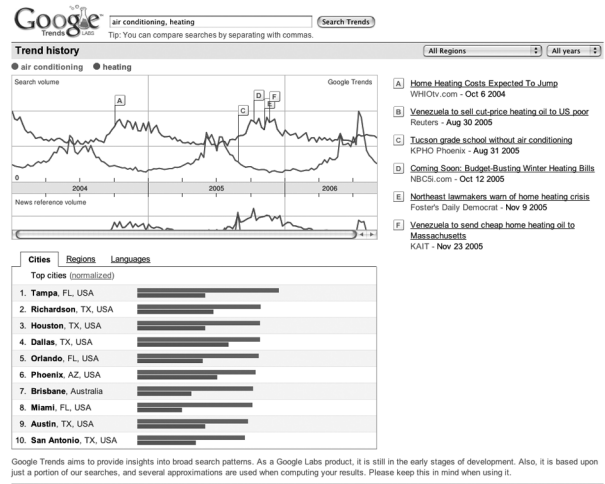
Aide utilisateur
Durée de la visite
 Combien de temps les visiteurs consacrent-ils à la visite de mon site Web ? Les visites dont la durée est inférieure à dix secondes apparaissent à gauche du graphique ; les visites dont la durée est supérieure à 1801 secondes apparaissent à droite du graphique.
 Pour comparer deux périodes, cliquez sur l'icône [] dans la barre de titre Période.

Un paradoxe réside dans la durée particulièrement rapide de 70% des visites, moins de 10 secondes. L'image mentale générée par le cerveau demande moins d'une seconde, ceci explique la rapidité de la lecture. Il ne faut pas compromettre cette lisibilité par un retard trop long dans l'affichage du contenu. L'optimisation de la vitesse pour un site Web reste d'actualité, même avec l'utilisation de connexion ADSL. Le lecteur intéressé peut parcourir mon article sur le sujet: FI10/05, dit.epfl.ch/publications-spip/article.php3?id_article=989.

TENDANCE CHEZ GOOGLE

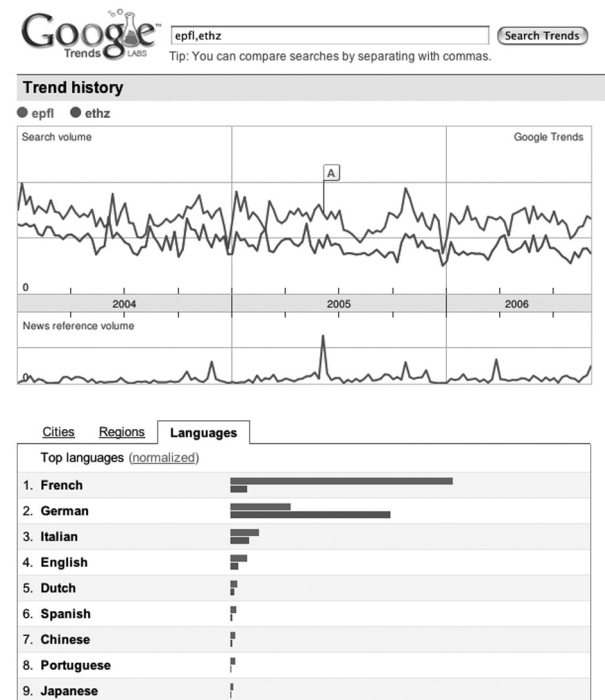
Si l'analyse détaillée présente des éclairages nouveaux et permet de confirmer ou d'infirmer ses intuitions, il peut être intéressant d'aborder le rivage d'une vision planétaire qui offre également quelques surprises!

Google Trends compare des mots-clés en fonction de leur demande passée dans le moteur de recherche: <http://www.google.com/trends>. L'anglais étant la langue véhiculaire du Web, il est préférable de l'utiliser dans les recherches, par exemple comparer l'**air conditionné** avec le **chauffage**: <http://www.google.com/trends?q=air+conditioning,+heating>.



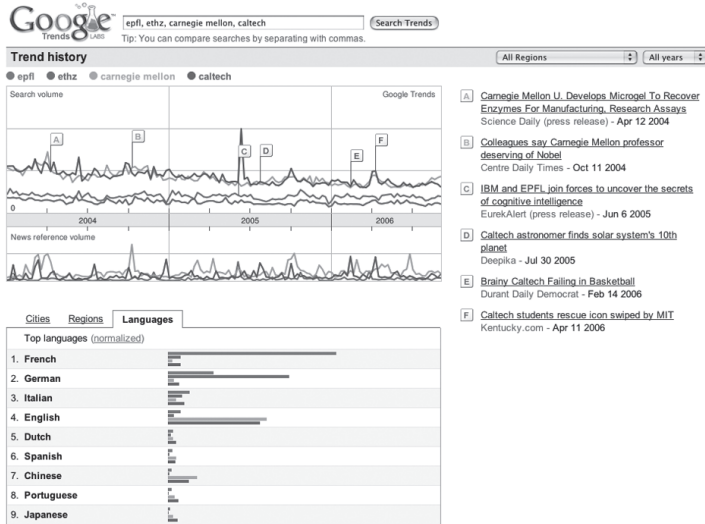
On observe les pics saisonniers. La base de données des demandes couvre approximativement 2 ans 1/2 et met en relation certains pics avec l'actualité.

On ne peut résister à comparer notre institution avec sa grande sœur de Zurich: <http://www.google.com/trends?q=+epfl,+ethz>.



Apparemment, notre institution est plus recherchée...

La demande d'information supplémentaire pour l'EPFL provient de l'extérieur. Comparons-nous avec quelques acteurs américains: epfl, ethz, carnegie mellon, caltech.



Google Trends aims to provide insights into broad search patterns. As a Google Labs product, it is still in the early stages of development. Also, it is based upon just a portion of our searches, and several approximations are used when computing your results. Please keep this in mind when using it.

©2006 Google - Google Labs - Discuss - Terms of Use - Privacy Policy - About Google Trends

Les courageux peuvent rajouter le MIT et Standford: epfl, carnegie mellon, caltech, mit, standford.

QUELQUES TENDANCES

prénoms	Laurent, Claudia, Jonathan, Katharina, Nicolas
Créationniste ou évolutionniste	intelligent design, evolution
Console de jeu	Wii, ps3, xbox 360
Bien ou mal	good, evil dieu, diable
Sécurité ou jeu	network security, tetris

CONCLUSION

On peut s'interroger sur le modèle de développement de logiciel classique, à la *Microsoft*, et leurs cortèges de correction face au modèle actuel d'applications disponibles par l'intermédiaire du Web. Dans le cas de **Google**, le nombre de services disponibles est saisissant, les internautes intéressés peuvent parcourir deux pages Web: www.google.com/intl/en/options/ et labs.google.com/.

Ces deux nouveaux outils gratuits fournis par **Google** permettent de mieux comprendre l'attrait du Web, pour un site ou sur un plan plus général. Ainsi, la compréhension du public réel offre un service de meilleure qualité. ■