

# Chronique

de la recherche sur le web

par Marc Duval

24 septembre 2002, vol.2, n°8

## Northern Light: La lumière du web invisible

**N**orthern Light est apparu sur le web en août 1997. Son nom a été emprunté à celui d'un clipper construit en 1851 à Boston. Il était réputé pour sa vitesse issue d'une nouvelle technique de construction. Contrairement au célèbre voilier, la venue de Northern Light est passée inaperçue pour le grand public bien qu'il ait été un précurseur dans plusieurs aspects de la recherche. Ainsi, il a été le premier à offrir la recherche dans les actualités et la catégorisation par dossiers du résultat d'une recherche avant AlltheWeb et Teoma. Il a apporté la recherche de localisation avant Google avec GeoSearch, fonction maintenant inactive. Il est le seul site de recherche d'importance à avoir ciblé sa clientèle, celle d'affaires, alors que tous les autres étaient généralistes. Enfin, il a été et demeure le premier site de recherche à offrir le web improprement qualifié d'invisible avec ses banques de données d'affaires.

2002 est une année charnière pour

Northern Light au même titre qu'elle le deviendra pour Inktomi.

Le 8 janvier, Northern Light annonce sa nouvelle orientation d'affaires. Désormais, la banque de données web n'est plus mise à jour pour le public mais reste ouverte pour la consultation. Son formulaire de soumission de pages web est fermé. Sa formidable collection de documents demeure active moyennant paiement pour la majorité d'entre eux.

Le 22 janvier, la société divine inc. achète Northern Light. Le lendemain, la collection de documents à accès payant est disponible sur Yahoo!. Ce nouveau service porte le nom de Yahoo! Premium Collection Search. Huit mois plus tard, le 24 septembre, le service est retiré.

Northern Light a cinq faiblesses et cinq forces.

Ces faiblesses sont:

1. l'accès difficile à ses formulaires de recherche
2. la conception de ses formulaires
3. la fermeture de sa banque de données web
4. la restriction de ses champs de recherche
5. la configuration des mots.

Ses forces sont:

1. un langage de recherche évolué
2. la fonction de tri
3. la catégorisation des références
4. la collection de banques de données
5. les alertes de recherche.

A ces forces et ces faiblesses, il faut ajouter une particularité qu'il met de l'avant: la recherche en langage naturel. En fait, le logiciel de recherche ne fera que retirer les mots de liaisons - appelés aussi mots vides - pour garder les mots significatifs. Il ne fait aucune interprétation de la requête que sous-tend la recherche en langage naturel. Cette caractéristique le différencie d'AltaVista qui recherche tous les mots d'une requête.

Soulignons aussi les conseils de recherche alternant au-dessus de la case de recherche de la page d'accueil.

## **Les faiblesses de Northern Light**

### **1. *L'accès au formulaire de recherche***

La page d'accueil de Northern Light est sous l'adresse <http://www.northernlight.com>. Elle a toutes les apparences d'une page habituelle d'un site de recherche. En y regardant de plus près, Northern Light ou NL donne l'accès à quatre formulaires détaillés: Power Search, Business Search, Search News et Investext Search. La case de recherche est accompagnée d'une liste déroulante de neuf banques de données: Special Collection, Investext Reports, MarkIntel Market Research, Archived News, American Banker, BusinessWeek, The Economist, Wall Street

Journal Absts, World Media Absts: Europe. Plus bas, un tableau donne les quatre services de divine, inc. : Special Collection, RivalEye (répertoire de liens), les services de recherche pour entreprises et les actualités. De recherche web, rien! Pour y avoir accès, il faut sélectionner le lien [NL Research](#) au bas de la page.

### **2. *La banque de données web***

NL était reconnu pour son haut niveau d'innovation. La fermeture de sa banque de données web est la principale perte qu'a connue le web. Sa page de soumission publique est maintenant fermée. Heureusement, nous avons encore accès à la recherche web, mais sa banque de données est devenue un «entrepôt d'archives».

### **3. *Les formulaires de recherche***

Le changement qu'a subi Northern Light en janvier 2002 n'a pas été pour le mieux. Au lieu de garder ou d'adapter les formulaires d'origine, la société divine les a doublés. Ainsi, les formulaires de Power Search et de Business Search de la page d'accueil n'ont pas la recherche web, contrairement à ceux de la page de NL Research. Ce n'est pas étonnant que plusieurs aient affirmé que la recherche web avait été fermée.

### **4. *Les champs de recherche***

Les formulaires détaillés sont minimaux. Il n'y a aucune case pour les opérateurs booléens, seulement la recherche générale et

la recherche par mots du titre et quelques spécificités propres à chaque catalogue. Dans Power Search, il y a cinq champs supplémentaires: la recherche par langue, par mots dans l'URL, par domaines, par intervalles de dates et par sujets comme «Tout le web francophone» de Voila.

Afin de faire une recherche optimale, l'internaute a besoin de connaître le langage et la syntaxe d'une requête évoluée comprenant l'usage des opérateurs booléens, des parenthèses et des quelques champs mis à notre disposition. NL le reconnaît en insérant des lignes d'aides aléatoires au-dessus de la case de recherche de la page d'accueil de NL Search.

#### 5. **Casse, diacritiques et dictionnaire**

Northern Light ne fait aucune distinction entre les lettres majuscules et minuscules si ce n'est une variation mineure dans le classement. Cette uniformité de forme ne permet pas de distinguer l'unité de mesure «pascal» du philosophe français Pascal, à tout le moins, de réduire le bruit.

L'absence de discrimination de mots à l'aide des signes diacritiques amène aussi son lot de bruit. Certainement pratique pour rechercher des variantes orthographiques comme « événement » et « évènement », mais embarrassant pour les mots tels que « ferme » et « fermé ».

Enfin, Northern Light ne contient aucun dictionnaire ou vérificateur orthographique, ni de suggestions de recherche qui sont devenus des outils incontournables.

### **Les forces de Northern Light**

#### 1. ***Le Langage de recherche évolué***

Northern Light n'a pas de formulaire détaillé comprenant les opérateurs booléens explicites. Ceci n'empêche pas que NL ait un langage de recherche tout à fait convenable pouvant s'exprimer dans une syntaxe idoine. NL a les trois opérateurs logiques conventionnels - AND (+), OR et NOT (-) - et l'opérateur de proximité « " " ». Malheureusement, il n'a pas l'opérateur d'adjacence permettant la recherche d'un couple de mots séparés par un mot et plus. Cependant, il a deux atouts: les parenthèses et la troncature. Les parenthèses permettent l'organisation syntaxique d'une requête complexe.

#### **Exemple:**

“opérateur booléen” AND (“moteur de recherche” OR “automate de recherche”) NOT troncature.

La troncature chez NL est la plus évoluée des automates de recherche. Elle se trouve sous trois formes: la troncature simple, la troncature à droite et la troncature interne. La troncature simple est par défaut: tous les mots formés avec le pluriel régulier sont recherchés en utilisant le singulier. La troncature droite, dont le symbole est l'astérisque « \* », est utilisée pour la recherche de toutes les déclinaisons d'un mot tandis que la troncature interne - % - remplace une lettre dans un mot. Il est à noter que la troncature interne est fonctionnelle après la quatrième lettre.

**Exemple:**

moteur ==> recherchera aussi moteurs  
 chercheu\* ==> recherchera «chercheur,  
 chercheurs, chercheuse et chercheuses»  
 occur%ence ==> recherchera « occurrence»  
 (faute courante) et « occurrence»

La fonction de troncature interne est utilisée pour la recherche de mot fautif et aussi d'orthographe voisine. Comme Northern Light s'adresse aux anglophones, cette fonction est particulièrement appréciée pour rechercher les mots américains et britanniques ayant des graphies différentes comme «analyse» et «analyze»

**2. La fonction de tri.**

NL Search est avec MSN Search les deux seuls automates qui appliquent le tri au résultat de recherche web. Les références peuvent être triées par pertinence et par date. Il a même une commande de recherche qui peut être ajoutée à une requête: SORT:date. Cette commande trie les références par ordre descendant.

**3. Catégorisation des références**

La catégorisation des références a été la marque de commerce de Northern Light. Il a été le premier à l'introduire dans un résultat de recherche. La catégorisation est faite à la volée. Elle s'inscrit à gauche du résultat sous la forme de dossiers. Elle permet l'affinage de la recherche.

**4. Collection de banques de données**

A ce jour, Northern Light constitue l'essence même du web dit invisible. Il est l'un des sites d'informations de langue anglaise les plus accessibles comparativement à d'autres fournisseurs comme Factiva. Sa banque de données web comprend plus de 350 millions de pages et continue de croître pour les entreprises abonnées à son service. Il continue de diffuser les actualités de sa collection et aussi de l'Associated Press dont la page comprend un logiciel de recherche pour des archives de 2 semaines. Il maintient aussi un répertoire de liens appelé RivalEye ou Special Editions. Il compte plus de 70 millions de textes tarifés répartis dans plus de 7 100 sources regroupées dans sa Special Collection ou Premium Collection. A cette collection, s'ajoutent les banques de données Investext Reports, World Media Absts: Europe, Wall Street Journal Absts, Business Week, MarkIntel Market Research Reports, Newspapers, Wires and Transcripts, City regional newspapers, College newspapers et Broadcast News Transcripts. Les rapports qui ont une valeur de milliers de dollars sont vendus à la page. Cette vente à la pièce permet de se procurer uniquement la partie qui vous intéresse. Pour un montant de 1\$US à 4\$US la page, vous pouvez avoir une information autrement inaccessible si vous aviez à vous abonner à un service avec frais de gestion comme le service de [Dialog](#). Un chapitre de rapport peut vous revenir à 20\$US. Enfin, certaines informations de presse peuvent être gratuites en utilisant Northern Light comme banque de données bibliographiques. La référence prise, il ne vous

reste qu'à vous diriger vers le site web de la revue ou du journal qui a peut-être mis l'article à la disposition des internautes.

Afin de faciliter la recherche dans ses banques de données, NL a ajouté trois champs de recherche qui leur sont propres: COMPANY: (pour la recherche par nom d'entreprise), PUB: (pour la recherche d'une publication), RECID: (pour le numéro identifiant une publication), TICKER:(pour la recherche par symbole boursier d'entreprise).

## 5. **Les alertes de recherche**

L'alerte de recherche est une requête automatisée diffusée sur un tableau d'affichage électronique ou par courriel. Bien que la banque de données web soit fermée, l'alerte chez Northern Light est encore active pour ses autres services, notamment les actualités. L'usage de l'alerte demande un enregistrement gratuit à partir du lien « My alerts». Cet enregistrement ouvre la porte à tous les services.

Le fonctionnement est simple. Lors d'une recherche, le lien « Save this Search as an Alert» est affiché en haut de la page du résultat. Ce lien vous amène à la page de sauvegarde de l'alerte dans laquelle la recherche est nommée et la source identifiée. La fonction de prévisualisation donne un résultat de recherche pour vérifier la pertinence de la stratégie et la modifier si c'est nécessaire. Si le résultat est convenable, la stratégie est sauvegardée en sélectionnant le bouton «Save alert». Une fois la stratégie sauvegardée, vous

pouvez aller sur votre page privée pour gérer vos alertes en les modifiant ou les supprimant. Dorénavant, vous recevrez par courriel un message vous avertissant de l'arrivée de nouveaux documents pour chacune de vos alertes. Ce message comprendra un lien vers le résultat de recherche sur le site de Northern Light.

## Conclusion

Northern Light a été le négligé des sites de recherche depuis sa parution sur la toile. Dès le début, il a été dans l'ombre d'AltaVista, d'Exite et de HotBot. Puis Google est arrivé, balayant tout sur son passage. Il a peut-être été une des raisons qui a motivé l'interruption de l'aventure de recherche web pour NL. Sous estimé hier, NL l'est encore aujourd'hui. Spécialisé dans le secteur des affaires, Northern Light demeure un site de premier ordre pour ses « bibliothèques » d'actualités, de rapports et de liens. Bien qu'il soit essentiellement tourné vers le marché de langue anglaise et sa banque de données web close, il ne peut être écarté dans une démarche de recherche pour des documents difficilement accessible par leur prix, des documents du « web invisible ».

---

Marc Duval

Service de recherche documentaire DSI  
1280-4, bd de Montarville.  
Longueuil, Québec, Canada  
J4B 8B4  
[marcduval@dsi-info.ca](mailto:marcduval@dsi-info.ca)  
<http://www.dsi-info.ca>