

Minoussink : le nouvel algorithme de Yandex qui a secoué les SEO russes



Par Philippe Yonnet
et Alexis Rylko

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Yandex est depuis des années leader des moteurs de recherche en Russie grâce notamment à des algorithmes différents de ceux de Google, prenant mieux en compte la langue russe. Mais cela ne l'empêche pas d'être confronté à des problèmes de spam, qu'il gère grâce à des outils et algorithmes spécifiques. Cet article a pour ambition de mieux vous faire connaître ce moteur très intéressant, et de vous parler de Minoussink, son dernier filtre qui lutte contre les liens de faible qualité.

La Russie fait partie de ces pays où la domination de Google que nous connaissons en Europe de l'Ouest n'existe pas. Un moteur résiste encore et toujours à l'envahisseur : Yandex. Est-ce la conséquence d'un protectionnisme local ? Sans doute un peu. Mais la vraie raison est que Yandex a réussi pendant des années à fournir de bien meilleurs résultats que Google pour la langue russe. Voilà donc un vrai champion local, doté d'un algorithme fondamentalement différent, qui dame le pion au colosse de Mountain View sur son terrain de prédilection : les algorithmes.

Mais la bataille se joue également sur plusieurs fronts. L'écosystème Internet Russe (le « RuNet ») est doté d'une tradition de linkspam et d'achats de liens massifs et industrialisés. Les principaux réseaux et sociétés intermédiaires ou de services dans ce domaine sont russes. Et la pertinence des résultats était grandement menacée par les pratiques massives de manipulation de liens. Nous allons donc voir comment Yandex essaie depuis

Yandex

quelque temps de se défendre contre cette menace, et en particulier, ce qu'a changé la dernière mise à jour de son algorithme, nommée « Minoussinsk ».

Yandex : le moteur dominant en Russie et un acteur avec lequel Google doit compter

Le moteur de recherche Yandex maintient une part de voix dominante en Russie avec plus de 58% du trafic généré par les moteurs de recherche. Mais Yandex est aussi un portail, et propose, comme Google, de nombreux autres services. La société Yandex et ses produits dominent le « search » en Russie, et Yandex est aujourd'hui clairement l'un des premiers media internet par son audience en Russie.

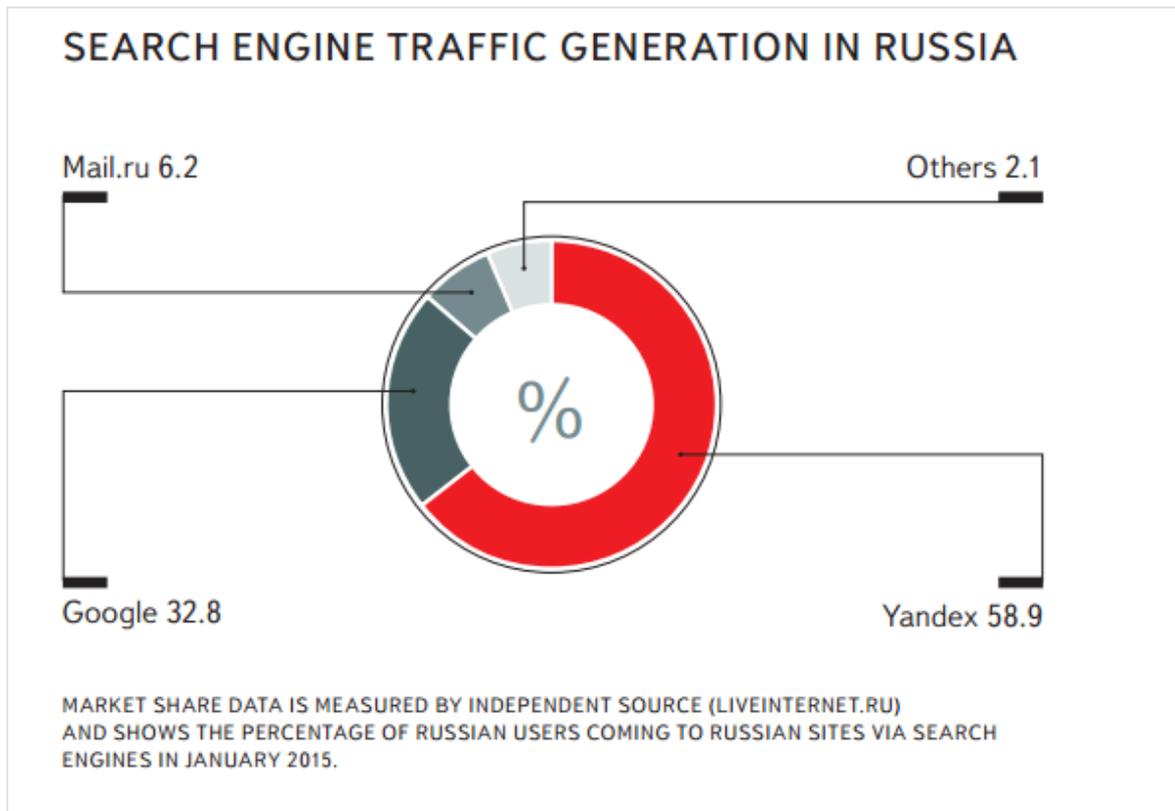


Fig.1. Parts de trafic des moteurs de recherche en Russie.

Sur le plan des fonctionnalités, le moteur Russe n'a pas grand-chose à envier à Google. Il existe beaucoup de points communs entre les deux moteurs et leur stratégie de services est similaire. Voici quelques exemples de services équivalents :

- Google Webmaster Tools (Search Console) -> Yandex Webmaster Tools ;
- Google Analytics -> l'excellent Yandex Metrica ;
- Google Adwords Suggestion Tool -> Yandex Wordstat ;
- Google Chrome -> Yandex Browser ;
- Etc.

Ne croyez pas pour autant que Yandex n'est qu'une pâle copie de Google : Yandex a souvent lancé certains services AVANT Google, et le moteur est en fait bien plus ancien que Google lui-même. Yandex propose également des services que Google ne propose pas ou plus

(comme par exemple Yandex Meteo, ou Yandex Catalogue, l'annuaire incontournable en Russie).

Un peu d'histoire

Les prémices du moteur remontent aux années 80, lorsque les deux principaux fondateurs du moteur, Ilya Segalovich et Arkady Volozh (voir photo ci-contre) qui travaillaient à l'époque pour la société Arkadia, ont développé les premières briques d'un outil de recherche adapté à la langue Russe. Ces recherches entraient dans le cadre d'un programme de l'Académie des Sciences de l'URSS, pour le Gouvernement de l'Union Soviétique.



The screenshot shows the Yandex search engine interface. At the top, the Yandex logo is on the left, and a search bar contains the text 'maria sharapova' with a subtext '— 2 million answers'. To the right of the search bar are icons for a close button, a filter icon, and a 'Search' button. Below the search bar, the results are categorized into 'Web', 'Images', 'Video', 'Translate', and 'More'. The 'Web' section features a link to 'Mariasharapova.com' with a subtext 'The official site with photos, videos, results, biographical information, articles and interviews.' The 'Video' section features a link to 'Maria Sharapova - Wikipedia, the free encyclopedia' with a subtext 'en.wikipedia.org > Maria Sharapova'. Below this, there is a section for 'Yandex.Video: 182 thousand videos found' with a grid of four video thumbnails. The first thumbnail shows Maria Sharapova on a tennis court with a subtext 'Maria Sharapova - Tokyo 2011 qtr youtube.com'. The second thumbnail shows Maria Sharapova with a subtext 'Maria Sharapova & #x27;s "injury"... youtube.com'. The third thumbnail shows Maria Sharapova with a subtext 'Maria Sharapova High Intensity Training (HIT) youtube.com'. The fourth thumbnail shows Maria Sharapova with a subtext 'Maria Sharapova vs Shahar Peer 2008 Fed Cup... youtube.com'. Below the video thumbnails, there is a section for 'Maria Sharapova (@MariaSharapova) | Twitter' with a subtext 'The latest Tweets from Maria Sharapova (@MariaSharapova). The Official Twitter Account for Maria Sharapova.' At the bottom, there is a section for 'See images based on search "maria sharapova"' with a grid of four image thumbnails. The first thumbnail shows Maria Sharapova sitting on a tennis court with a subtext 'images.yandex.com > maria sharapova'. The second thumbnail shows Maria Sharapova in a bikini. The third thumbnail shows Maria Sharapova in a white tank top. The fourth thumbnail shows a close-up of Maria Sharapova with a subtext 'More images >'.

Fig.2. Une page de résultats de Yandex version anglophone.

Dès 1993, ils choisissent de baptiser ce moteur « Yandex », pour « Yet ANother inDEX ».

Le moteur de recherche grand public Yandex.ru a été lancé le 23 septembre 1997, soit un an avant Google ! Et les premières publicités contextualisées apparaissent sur les pages de résultats du moteur en 1998, bien avant Adwords.

Arkady Volozh a ensuite créé une société en 2000, avec l'aide de plusieurs investisseurs, pour « porter » le moteur et son activité. Il en est le CEO depuis lors. La société Yandex est maintenant propriétaire du premier media Internet de Russie, est entrée au NASDAQ en 2011, et a entamé une

expansion internationale, notamment en posant le pied de manière remarquée en Turquie.

Un algorithme fondamentalement différent de celui de Google

Depuis le début, les créateurs de Yandex ont suivi un chemin différent du paradigme dominant de la recherche aux Etats Unis.

L'adaptation aux langues flexionnelles et la détection de l'intention

Le Russe est une langue fortement flexionnelle, c'est-à-dire que les mots changent de forme en fonction de leur cas, de leur genre, de leur nombre, de la conjugaison du verbe etc... Cette caractéristique atteint même les noms propres (Mme Poutine s'appelle « Putina » M. Poutine : « Putin »). Une autre spécificité de la langue russe par rapport à l'anglais est que l'ordre des mots n'a aucune importance dans cette langue.

Ces caractéristiques sont une malédiction quand l'algorithme n'est pas fait pour reconnaître les flexions et se base sur une grammaire différente. Altavista, Google, ou Msn ont d'abord été conçus pour l'anglais, langue qui connaît peu le phénomène des flexions, et dans laquelle l'ordre des mots aide à comprendre le sens de la phrase. Ceci explique pourquoi Yandex, qui dès le départ a été conçu pour gérer la langue russe, s'est taillé durablement la part du lion.

Très tôt, les concepteurs de l'algorithme ont également fait des choix différents sur bien d'autres aspects : au lieu d'essayer de mesurer une similarité entre la requête et les documents, ils ont essayé de mesurer la probabilité qu'un document soit pertinent sur une requête donnée. Et cette probabilité dépend de la détection de l'intention cachée de l'internaute derrière la saisie des mots clés. Pour détecter cette intention, la reconnaissance des flexions est souvent cette fois une bénédiction.

Enfin, contrairement à Google qui s'est longtemps refusé à cette approche, jugée peu scalable et inadaptée au

volume gigantesque de données qu'ils traitaient, Yandex s'est lancé très tôt dans l'exploitation de l'apprentissage automatique pour améliorer ses pages de résultat.

Matrixnet ou le machine learning à tous les étages

En 2009, Yandex annonce un saut quantique dans l'exploitation des algorithmes d'apprentissage automatique en lançant Matrixnet, une plateforme logicielle capable d'apprendre comment exploiter les multiples « signaux » (les critères) que gère le moteur pour classer correctement les pages selon de nombreux types de requêtes. Selon Yandex, Matrixnet apporte surtout une avancée dans la résolution d'un problème baptisé « l'overfitting » (le surapprentissage). Un algorithme de moteur basé sur l'apprentissage automatique « invente » lui-même la « formule » qui combine tous les signaux pour classer les résultats par ordre de pertinence. En cas de surapprentissage, la formule peut être ridiculement compliquée, rendant son exploitation impossible sur un moteur qui doit fonctionner en temps réel.

La plateforme Matrixnet a aussi permis à Yandex de changer son approche de la lutte contre le spam et le rôle des « quality raters ». Son déploiement a permis de développer de nouveaux algorithmes capables d'éliminer de gros volumes de spam à partir des données d'entraînement entrées par les quality raters.

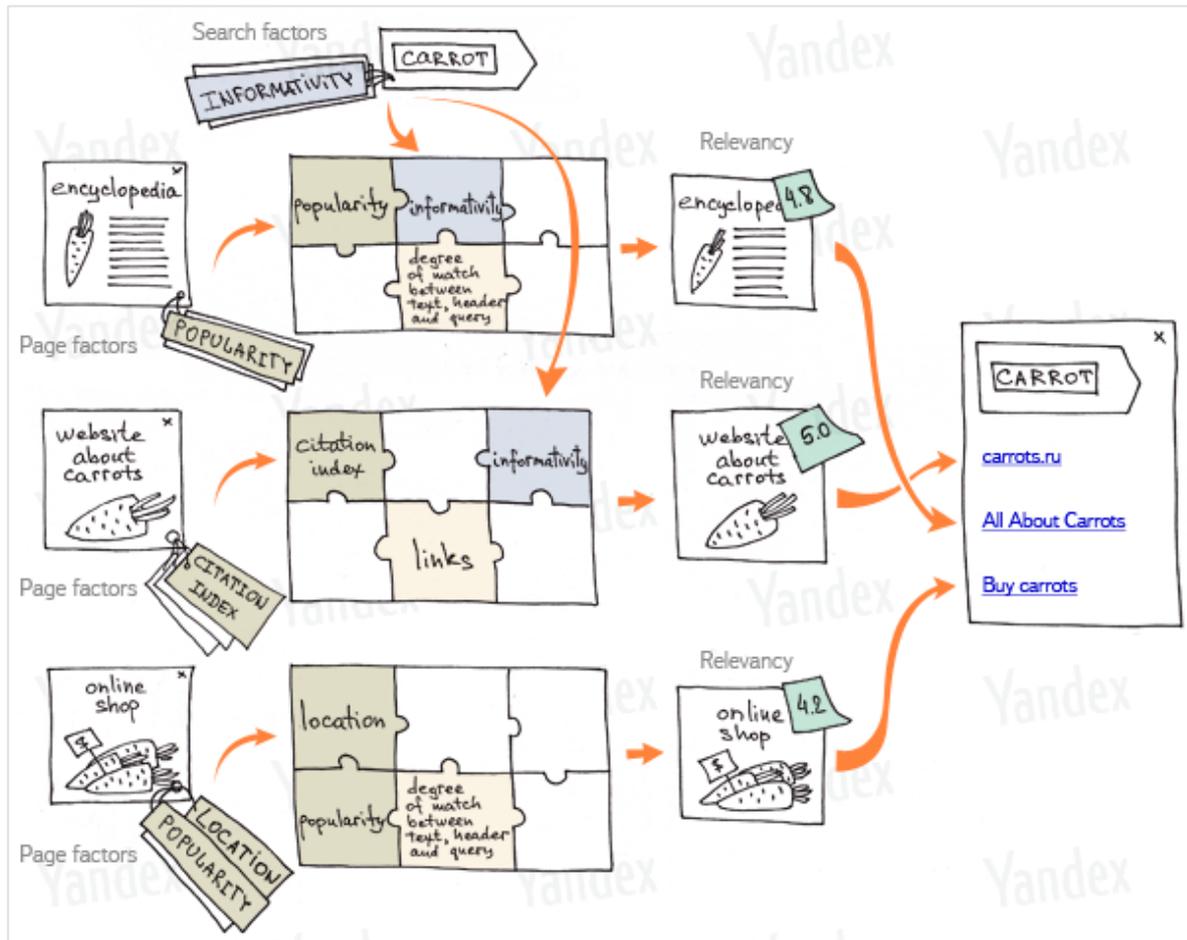


Fig.3. Les différents groupes de signaux utilisés par Yandex.

Les principaux critères utilisés par le moteur Yandex

La plupart des critères utilisés sont communs à d'autres moteurs comme Bing ou Google :

- La pertinence textuelle (correspondance entre le contenu de la page et la requête) ;
- La popularité (backlinks) : Yandex utilise un indice différent du Pagerank baptisé le TIC ;
- Les critères techniques.

Yandex a une approche spécifique pour 3 familles de critères :

- L'expérience utilisateur : ces critères sont importants pour l'algorithme ;

- La correspondance régionale : l'approche est adaptée à la géographie de la Russie ;
- La pertinence commerciale : Yandex identifie les requêtes commerciales et cherche à renvoyer des pages e-commerce de qualité en évaluant les pages des sites e-commerce en fonction de critères de qualité spécifiques.

Les principaux types de pénalités sur Yandex

Yandex dispose comme tous les moteurs d'un système de pénalités infligées aux sites qui violent ses guidelines. Voici les plus connues de ces pénalités :

- Pénalité pour emploi de pages satellites => disparition des résultats ;
- Cloaking => disparition des résultats ;
- Texte caché => disparition des résultats ;
- Contenu de faible qualité (pénalité AGS équivalent de Panda+Penguin) => disparition totale ou partielle du site dans les résultats, recul important de positions, baisse du TIC ;
- Thin affiliates (sites affiliés sans valeur ajoutée) => disparition totale ou partielle du site dans les résultats, recul important de positions ;
- Suroptimisation (ex *keyword stuffing*) => disparition totale ou partielle du site dans les résultats, recul important de positions ;
- Manipulation des critères d'usage (critères compartementaux) => disparition totale ou partielle du site dans les résultats, recul important de positions ;
- Publicité excessive ou intrusive => recul important de positions.

Mais le principal problème auquel Yandex est confronté reste la manipulation des backlinks par les SEO : il a donc récemment introduit de nouvelles pénalités pour linkspam.

Yandex et le linkspam

Pendant longtemps, la politique de Yandex par rapport à l'achat de liens, le linkspam ou la création de galaxie de sites pour créer des backlinks, a été de détecter les liens manipulés et d'annuler leur impact sur l'algorithme. Mais Yandex, comme tous les moteurs, s'est rendu compte que sa détection des liens manipulés était perfectible et qu'il fallait adopter une autre approche.

Le 6 décembre 2013, à la conférence IBC Russia, **Yandex annonce l'abandon de liens** l'année prochaine dans le classement sur les requêtes commerciales. Raison principale : suite à des tentatives d'achat massif, les backlinks ont perdu leur valeur. Heureusement il existe encore 800 facteurs de classement. Le 12 mai 2014, la confirmation tombe : à la conférence Bynetweek à Minsk, Alexandre Sadovsky de Yandex Search annonce l'abandon de la prise en compte de liens entrants dans la région de Moscou dans les thématiques suivantes :

- Immobilier ;
- Tourisme ;
- Electronique et outils ménagers ;
- Vêtements et accessoires ;
- Beauté et santé ;
- Services juridiques ;
- Création de site et SEO.

Les avis de la communauté SEO russophone sont partagés, et certains doutent de la sincérité de l'annonce. Le principal problème étant que Yandex, à part son annonce spectaculaire, n'a présenté aucune preuve que les backlinks aient été réellement retirés de l'algorithme de classement.

Le 15 avril 2014, Yandex déploie une mise à jour du filtre antispam "AGS" visant les sites vendant des liens et ceux de faible qualité. Dorénavant au lieu de désindexer ces sites, Yandex va réduire leur indice TIC (Indice de citation thématique) à zéro.

L'annonce de l'abandon de liens a tourné les SEO vers d'autres leviers permettant d'influencer les classements, en particulier manipuler les critères liés à l'expérience utilisateur (en simulant une expérience utilisateur correcte).

Plusieurs outils en ligne ont vu le jour et ont lancé une tendance inédite.

Mais le 13 mai 2014, Yandex lance un algorithme antispam pénalisant les sites recourant à ces techniques de manipulation des critères liés à l'UX.

Yandex a signalé que la seule solution pour enlever cette pénalité est d'arrêter d'imiter l'expérience utilisateur. Mais vu que ces techniques présentent une violation grave des guidelines de Yandex, les sites pénalisés se verront infliger une perte de visibilité d'une durée de 6 mois !

La gravité de cette pénalité nous apprend du coup quelque chose : les moteurs de recherche n'introduisent pas de pénalités pour les choses qui n'influencent pas leurs résultats de recherche, donc l'expérience d'utilisateur est un facteur de classement important pour Yandex

Changement d'approche en 2015 : retour (implicite) des backlinks dans l'algorithme et pénalités pour achat de liens

Le 15 avril 2015, à la conférence Bynetweek à Minsk Alexandre Sadovsky, responsable de l'équipe anti-spam de Yandex, annonce la mise en place de pénalités visant les acheteurs de liens. Le même jour l'équipe de Yandex Search envoie 9000 messages via son outil pour webmasters avec le même contenu : « ...Parmi les sites associés à votre compte, nous avons trouvé des sites qui utilisaient dans leur référencement un grand nombre de backlinks SEO... Afin d'éviter des pertes de classement, nous recommandons fortement de renoncer à l'utilisation de

liens SEO et d'autres méthodes inappropriées de référencement. ».

Donc, contrairement aux pénalités d'autrefois, visant principalement les vendeurs de backlinks, cette fois-ci ce sont des acheteurs qui sont dans le collimateur.

Minoussinsk : nuit du 15 au 16 mai 2015

Le déploiement du filtre « anti achat de liens » est intervenu dans la nuit du 15 au 16 mai 2015. Il a été baptisé Minoussinsk (les mises à jours de Yandex se voient attribuer des noms de localité).

La première itération de pénalité "Minoussinsk" atteint 488 sites, leaders dans les volumes d'achat de liens. Tous les sites perdent de 14 à 25 positions sur les requêtes phares.

A côté de la pénalité purement algorithmique, on voit très clair une intervention manuelle avec le même effet. Elle concerne la thématique du "référencement naturel" dans laquelle Yandex a pénalisé quelques-unes des plus grosses agences SEO russes. La communauté a déjà appelé ce phénomène la "Fusillade publique".

Une deuxième vague de pénalités est attendue, qui impacterait beaucoup plus de sites dans plus de régions.

Le bilan de Minoussinsk

Yandex, a présenté un bilan du déploiement de Minoussinsk lors de la conférence Yandex Webmasterskaya (« Yandex Atelier ») le 21 mai 2015. En voici les grandes lignes :

1. 9000 messages d'avertissement ont été envoyés aux webmasters ;
2. 37% de sites ont réduit considérablement leur nombre de backlinks ;
3. 488 sites ont été atteints suite à la première itération de l'algorithme ;
4. Les itérations suivantes seront déployées sans envoi d'avertissement ;
5. Jusque-là, aucun site n'a rétabli ses positions ;
6. La part de backlinks SEO dans la base de Yandex a diminué de 21,7% ;
7. La part d'acheteurs de backlinks SEO dans le TOP-10 a diminué de 12,5% ;
8. Suite à une demande, l'équipe de support de Yandex Search donne une réponse directe si le site est pénalisé par Minoussinsk ;
9. Minoussinsk ne peut pas être soigné par le changement du nom de domaine. « Tôt ou tard Minoussinsk vous rattrapera » ;
10. Yandex ne prévoit pas de créer l'outil de désaveu de backlinks comme l'a fait Google.

Yandex est-il en crise ?

Ces récents atermoiements sur la question des backlinks et la succession rapide du déploiement des filtres antispam révèlent que le moteur cherche à améliorer à marche forcée la qualité de ses pages de résultats et d'influencer la communauté SEO russe en les détournant de certaines pratiques. Jusqu'ici, son avance technologique sur la langue russe l'avait protégé de la concurrence, mais il s'avère que Google a progressé dans ce domaine et commence à grignoter sérieusement les parts de marché de Yandex en Russie. Jugez plutôt : croissance de la part de voix de Google en Russie sur le trafic de recherche :

- Février 2015: 39.5% ;
- Mars 2015: 39.9% ;
- Avril 2015: 40.1% ;
- Mai 2015: 41.1%.

(Données publiques de Liveinternet: <http://www.liveinternet.ru/stat/ru/search.es.html?date=2015-05-01;period=month>)

Ce réveil plutôt tardif de Yandex dans la lutte contre le linkspam lui permettra-t-il de nettoyer ses pages et de garder son avance sur Google ? L'avenir (proche) nous le dira...

Bibliographie et liens

- Yandex en anglais <http://www.yandex.com>
- La présentation de Matrixnet en anglais : <https://company.yandex.com/technologies/matrixnet.xml>
- Les user agents de Yandex (ils sont nombreux) : <https://help.yandex.com/search/robots/logs.xml>
- Quelques ressources pour approfondir votre connaissance de Yandex : Russian Search Tips (en anglais), le blog d'Anna Oshkalo <http://www.russiansearchtips.com/>
- Le Blog d'Alexis Rylko (en français !) : <http://www.alekseo.com>
- Et son billet sur Minoussinsk : <http://alekseo.com/minoussinsk-algorithme-yandex/>



Philippe YONNET, Directeur Général de l'agence Search-Foresight, groupe My Media (<http://www.search-foresight.com>) et **Alexis RYLKO**, consultant SEO, Yandex et Google (Minsk – Biélorussie) (<http://www.alekseo.com>)