

Affaire suivie par :  
CERTA

## NOTE D'INFORMATION DU CERTA

### Objet : Outils d'indexation et de recherche

---

Conditions d'utilisation de ce document : <http://www.certa.ssi.gouv.fr/certa/apropos.html>  
Dernière version de ce document : <http://www.certa.ssi.gouv.fr/site/CERTA-2006-INF-009>

---

### Gestion du document

Référence	CERTA-2006-INF-009
Titre	Outils d'indexation et de recherche
Date de la première version	21 novembre 2006
Date de la dernière version	–
Source(s)	<a href="http://desktop.google.com">http://desktop.google.com</a>
Pièce(s) jointe(s)	Aucune

TAB. 1 – *Gestion du document*

Une gestion de version détaillée se trouve à la fin de ce document.

## 1 Introduction

Les logiciels d'indexation et de recherche se multiplient. L'élargissement de la couverture fonctionnelle s'accompagne d'une augmentation des risques de divulgation d'informations. La suite de la note d'information s'appuie sur un exemple, Google Desktop Search (GDS) version 3. Les recommandations s'appliquent également aux logiciels équivalents des autres éditeurs.

L'objectif de cette note d'information est d'identifier les menaces liées à l'utilisation de tels logiciels et de recommander de bonnes pratiques pour s'en protéger.

### 1.1 Fonctionnalités

Les logiciels d'indexation et de recherche proposent un affichage unifié des résultats de recherche, que les sources affichées soient sur Internet, sur l'ordinateur de l'utilisateur, voire sur plusieurs ordinateurs, sans que l'information soit préalablement structurée. Accessoirement, ces logiciels proposent la personnalisation du bureau et l'affichage d'informations en continu (météo, cours de bourse, actualités, état de l'ordinateur...).

Le périmètre d'indexation et de recherche sur un ordinateur comprend :

- des documents (texte, MS-office, Acrobat, pages HTML) ;
- des images, des fichiers musicaux et vidéo ;
- des courriels lus avec des clients POP/IMAP (Outlook Express, Thunderbird, Mozilla, etc.) ;
- des pages Web visitées, dont les courriels (webmail) ;

- des conversations (*chat*) par messagerie instantanée.

Les évolutions du logiciels et la possibilité d'extensions (*plug-in*) ne feront qu'allonger la liste des objets indexables.

## 1.2 Logiciels

Les logiciels existants ou émergents sont édités par Google (GDS), Microsoft (WDS), Yahoo, AOL, Copernic, AK Jeeves. La gratuité est de mise.

## 2 Installation et utilisation

Le logiciel est disponible gratuitement à l'adresse <http://desktop.google.com> (voir documentation). Il ne fonctionne actuellement que sous Windows 2000 SP3 ou XP.

Après l'installation et la configuration, le logiciel opère l'indexation, en tâche de fond, pour permettre les recherches, au gré de l'utilisateur.

### 2.1 Installation

L'installation, comme la désinstallation, ne peut s'effectuer qu'avec des droits d'administration.

La société Dell installe GDS sur les ordinateurs qu'elle livre depuis la signature d'un partenariat avec Google (voir documentation). L'installation par téléchargement, démarre effectivement après acceptation de la licence d'utilisation.

Une fois installé, le logiciel s'exécute pour tous les comptes et se lance à chaque démarrage de l'ordinateur. L'installation standard ajoute une icône sur le bureau et dans la barre de tâche. Techniquement, un serveur HTTP écoute sur le port 4664/TCP les requêtes de recherche en provenance de l'ordinateur local et les traite selon la configuration choisie par l'utilisateur.

### 2.2 Configuration

La configuration est faite par chaque utilisateur de l'ordinateur, sans nécessiter les droits d'administration.

Le lancement de GDS se fait en cliquant sur l'icône du bureau ou celle de la barre de tâches. Il permet la modification de la configuration par utilisateur (mémoire dans le profil Windows). Celle-ci se traduit par un choix de préférences, réparties en quatre onglets :

#### Indexation locale :

**Type de recherche** types de fichier indexés ;

**Indexation des plug-ins** téléchargement d'extensions ;

**Indexer et Ne pas indexer dans ces emplacements** répertoires, URL, lecteurs ;

**Crypter l'index** chiffrement de l'index avec le système EFS de Windows ;

**Désactiver l'indexation** mais n'agit pas rétroactivement ;

**Effacer les éléments supprimés** purge du cache.

#### Fonctionnalités de compte Google :

**Gmail** indexation des courriels d'une boîte Gmail ;

**Enregistrer le contenu et les paramètres de mon Google Gadget** accessibilité depuis différents ordinateurs ;

**Rechercher sur les différents ordinateurs** suppose l'activation des fonctionnalités avancées ;

#### Affichage dont :

**Intégration Google** recherche GDS intégrée à chaque recherche web.

#### Fonctionnalités avancées

activation on non.

La recherche par GDS sur plusieurs ordinateurs demande que :

- l'utilisateur dispose d'un compte Google ;
- sur chacun de ces ordinateurs, GDS soit installé ;
- le compte Google soit utilisé dans le paramétrage de GDS sur tous ces ordinateurs.

## 2.3 Indexation

Dès la configuration par l'utilisateur terminée, l'indexation commence. Elle s'effectue dès que l'ordinateur est inactif 30 secondes. L'index est créé dans le profil de l'utilisateur.

Dans le même temps, GDS copie chaque fichier indexé (constitution d'un cache). Tout changement sur un fichier (nouveau ou indexé) provoque la mise à jour de l'indexation et la copie de la nouvelle version ou du nouveau fichier dans le cache.

Sauf choix contraire de l'utilisateur, les documents MS-Office protégés par mot de passe et les pages web chiffrées (sites en HTTPS://) sont indexés et copiés.

L'option de recherche sur plusieurs ordinateurs (recherche multiordinateur) provoque la copie des données indexées sur les serveurs de Google. La protection du transfert du poste vers les serveurs Google repose sur l'utilisation de protocole HTTPS. Selon la politique actuelle de confidentialité de Google, ces informations sont conservées au plus 30 jours après leur dernière utilisation ou la clôture du compte Google.

## 2.4 Recherche

Toute recherche est traitée par le serveur HTTP local qui aiguille la requête en fonction de la configuration et de l'intégration choisie : poste local, web ou les deux. La présentation suit la configuration d'affichage.

La recherche multiordinateur utilise les données stockées sur les serveurs de Google. Elle peut viser des ordinateurs éteints ou déconnectés au moment de la recherche.

# 3 Menaces

Le fonctionnement des logiciels d'indexation et de recherche est susceptible d'ouvrir des brèches dans les systèmes d'information.

## 3.1 Divulgaration d'informations

GDS copie chaque fichier indexé, dans ses versions successives, et le conserve même après destruction du fichier original. Ceci s'applique également aux courriels, aux conversations (*chat*) et aux pages web visitées. Par conséquent les informations contenues dans un fichier, une page web ou un courriel restent accessibles après leur suppression du fichier et après suppression du fichier ou du courriel d'origine ou après déconnexion du site web.. La méconnaissance par les utilisateurs des fonctions de purge de GDS est susceptible de mettre en défaut le respect d'obligations de destruction d'informations, en particulier nominatives.

Si l'utilisateur n'a pas exclu du périmètre d'indexation les pages web chiffrées, les documents MS-Office protégés par mot de passe et les dossiers contenant des informations sensibles, l'indexation du cache couvre ces informations sensibles et les copie dans le cache.

Le logiciel GDS transmet à Google des informations sur les habitudes de navigation de l'utilisateur ayant opté pour les fonctions avancées. Ces informations sont exploitées pour fournir des fonctions supplémentaires (exemple : titres de l'actualité). La politique de confidentialité de Google n'exclut pas l'utilisation des données collectées à d'autres fins.

La copie de données sur les serveur de Google (recherche multiordinateur) augmente les risques :

- évocation d'informations sensibles (nominatives, financières, secrets industriels) ;
- fuite de critères de recherche ;
- cible centralisée pour la criminalité informatique.

Le partage d'un profil Windows par plusieurs utilisateur, non recommandé, se rencontre pour des postes accessibles au public (bornes, cybercafés). Il induit un partage d'index et de cache Google entre les utilisateurs du profil. Un nomade qui connecte un support amovible à un poste à profil partagé expose les informations sur le support à être divulgués via le cache. Les logiciels d'indexation ne font qu'allonger la liste des menaces auxquelles le nomadisme expose les informations transportées.

## 3.2 Consommation de ressources

La consommation du processeur est marginale : l'indexation ne s'effectue que lors de l'inactivité de l'ordinateur. Elle peut également être suspendue par l'utilisateur.

Le cache occupe un espace disque qui devient conséquent. La consommation est accrue par les modifications fréquentes des fichiers (copie des versions successives).

La recherche multiordinateur et l'indexation des disques réseau (ressources partagées) se traduisent par une occupation du réseau proportionnelle au volume d'informations traitées et peuvent encombrer significativement les réseaux. La première indexation est la plus coûteuse.

## 3.3 Intégrité du SI

L'indexation des courriels se traduit par l'ouverture de ceux-ci et de leurs pièces jointes. Cette opération peut provoquer l'exécution de virus de messagerie. Des virus pourraient être spécialement conçus pour exploiter une faille de GDS.

Les extensions (indexation de nouveaux types de fichiers, gadgets) sont développées à partir de l'API de Google et distribuables sans contrôle des codes développés. Elles représentent un vecteur d'introduction de logiciels instables ou destructeurs et de chevaux de Troie dans le système d'information.

## 3.4 Autres menaces

L'exploitation, par Google, des habitudes de navigation des utilisateurs, n'est pas limitée par la politique de confidentialité. Elle pourrait engendrer des flux indésirables, par exemple des fenêtres (*pop-up*) et des courriels publicitaires.

L'absence de contrôle sur la destruction effective des données stockées sur les serveurs Google n'est pas compatible avec un engagement ou une obligation de destruction de données.

Les serveurs collectant les données des utilisateurs ne sont pas hébergés dans l'Union européenne. Le régime juridique applicable à l'accès aux informations s'en ressent. D'une part, les utilisateurs de GDS ne bénéficient pas de l'application des lois françaises et des réglementations européennes les protégeant. D'autre part, des lois ou des procédures en vigueur dans les États où les serveurs sont hébergés peuvent autoriser l'accès des données à des tiers.

# 4 Recommandations

## Recommandation 1

N'installer Google Desktop Search que sur les postes dont les utilisateurs en ont l'utilité. Ne pas installer ce logiciel en standard.

## Recommandation 2

- Ne pas installer Google Desktop Search sur des postes ayant au moins l'une des caractéristiques suivantes :
- pratiques imparfaites en matière de sécurité (mises à jour du système et des applications non maîtrisées, absence d'antivirus) ;
  - dont plusieurs utilisateurs partagent un même profil (exemple : postes publics) ;
  - traitant des informations sensibles ;

## Recommandation 3

Lorsque Google Desktop Search est installé, ne pas activer par défaut :

- l'intégration de GDS ;
- l'utilisation multiordinateur ;
- l'enregistrement et le contenu du Google Gadget

- l'indexation des disques réseau ;
- l'indexation de messageries instantanées ;
- l'indexation des documents MS-Offices et des pages web sécurisés.

#### **Recommandation 4**

Centraliser la gestion du logiciel, par exemple en utilisant la version Entreprise dans un environnement Active Directory. Prévoir une désactivation centralisée, par exemple en cas de faille critique non corrigée.

#### **Recommandation 5**

Sensibiliser l'utilisateur de GDS aux risques auxquels il expose le système d'information et ses propres données personnelles, Le former sur les moyens de contrôler l'activité du logiciel, d'en modifier les options et de supprimer des informations du cache Google et sur les serveurs Google.

#### **Recommandation 6**

Responsabiliser l'utilisateur d'un support amovible hors du système d'information de l'organisme (le nomade). Il doit :

- être sensibilisé aux risques auxquels il expose les informations professionnelles et personnelles qu'il transporte sur ce support, de manière générale ;
- prendre toutes les précautions utiles pour que les informations transportées ne soient pas capturées par un logiciel fonctionnant sur un ordinateur auquel il connecte le support, en particulier par un logiciel d'indexation et de recherche.

### **5 Documentation**

- Site de Google Desktop Search :  
<http://desktop.google.fr/>
- Annonce du partenariat Dell-Google :  
<http://googleactu.blogspot.com/2006/05/google-dell-partenariat-confirm.html>
- Politique de confidentialité de Google :  
<http://www.google.com/privacy.html>

### **Gestion détaillée du document**

21 novembre 2006 version initiale.