

Google Search Appliance

Gouvernance du déploiement et modèles opérationnels

Mars 2014



© Google 2014

Gouvernance du déploiement et modèles opérationnels

Ce document présente quelques stratégies relatives à la gouvernance d'une solution Google Search Appliance (GSA).

À propos de ce document

Les recommandations et informations rassemblées dans ce document sont le résultat de notre travail sur le terrain avec de nombreux clients et partenaires dans des environnements variés. Nous tenons à les remercier chaleureusement d'avoir partagé avec nous leurs expériences et leurs observations.

| | |
|-----------------------------------|--|
| Thèmes abordés | Ce guide traite des bonnes pratiques en matière de publication, des stratégies de gouvernance des données, des processus embarqués relatifs au contenu et des modèles d'appartenance des solutions. |
| Lecteurs cibles | Architectes d'affaires, architectes de solutions, architectes techniques, administrateurs GSA et administrateurs de systèmes de gestion de contenu d'entreprise travaillant sur la conception d'une solution de recherche en entreprise basée sur Google Search Appliance. |
| Environnement informatique | GSA dans un environnement informatique centralisé, décentralisé ou tiers. |
| Phases de déploiement | Planification et déploiement de GSA |
| Autres ressources | <ul style="list-style-type: none">• Les thèmes abordés dans ce manuel sont expliqués de manière plus détaillée dans d'autres documents. Tous les documents GSA en pratique sont donc pertinents.• Le site Web LearnGSA.com fournit des ressources pédagogiques sur le système GSA.• La documentation produit de GSA fournit des informations complètes sur le système.• Le Portail d'assistance Google for Work permet d'accéder à l'assistance Google. |

Sommaire

[À propos de ce document](#)

[Chapitre 1 : Gouvernance de la publication de contenu et bonnes pratiques](#)

[Présentation](#)

[Évaluation de la pertinence et du classement](#)

[Création et structure du contenu](#)

[Options de réglage de la pertinence](#)

[Points à prendre en compte concernant l'expérience utilisateur](#)

[Chapitre 2 : Appartenance du système GSA et modèle d'imputation des coûts](#)

[Présentation](#)

[Appartenance à une organisation informatique centralisée](#)

[Appartenance à une seule unité fonctionnelle ou commerciale](#)

[Appartenance à un tiers dans un environnement hébergé](#)

[Gestion des opérations et assistance continues](#)

[Pilotage continu de la recherche](#)

[Chapitre 3 : Modèles d'acquisition de contenu de GSA](#)

[Présentation](#)

[Prise en compte des sources de contenu pour l'indexation](#)

[Comparaison entre Référentiel/plate-forme de contenu et modèle d'acquisition d'application d'entreprise](#)

[Processus continu d'acquisition et de maintenance](#)

[Gestion des données sensibles nouvellement indexées dans le système GSA](#)

[Chapitre 4 : Approche de l'environnement GSA](#)

[Présentation](#)

[Environnement de développement/Fonctionnalités expérimentales](#)

[Environnement de test et d'assurance qualité \(facultatif\)](#)

[Environnement de production](#)

[Récapitulatif](#)

Chapitre 1 : Gouvernance de la publication de contenu et bonnes pratiques

Présentation

Google Search Appliance fournit des algorithmes "prêts à l'emploi" permettant d'indexer et de classer avec efficacité les documents d'entreprise usuels. Les utilisateurs n'ont pas besoin de créer des documents répondant aux algorithmes de classement et d'indexation du système GSA, car le système GSA est capable d'indexer du contenu non structuré.

Il est néanmoins possible d'améliorer le système en fournissant au système GSA des signaux supplémentaires permettant d'améliorer le classement. Ces signaux incluent notamment l'emploi, dans les documents, de métadonnées appropriées (titre, auteur, date de publication) et de contenus intéressants. Plus il y a de signaux différenciateurs pris en compte par le système GSA, plus le classement de la pertinence du contenu est optimisé.

Évaluation de la pertinence et du classement

Google Search Appliance exploite de nombreux facteurs pour classer les résultats de recherche et fournit un niveau élevé de pertinence sans configuration supplémentaire. De nombreux algorithmes utilisés par le système de recherche ont été optimisés pour le contenu d'entreprise, ce qui permet à vos utilisateurs d'effectuer des recherches plus efficaces et de tirer davantage parti de votre corpus existant rassemblant les documents d'entreprise les plus utilisés (documents de traitement de texte, feuilles de calcul et présentations, par exemple).

Google Search Appliance fournit également des options de réglage intelligentes et intuitives, parmi lesquelles la pondération de nœuds, de collections, de la source, de la date et des métadonnées. En outre, le système de classement de GSA permet aux administrateurs de s'appuyer sur les journaux du serveur, ainsi que sur d'autres informations spécifiques à l'entreprise, pour affiner automatiquement le classement des résultats pour un nombre illimité d'URL. Google reste leader par la qualité et l'innovation de son système de recherche, grâce aux milliards de recherches effectuées sur la plate-forme et aux milliers d'ingénieurs voués à la technologie de recherche.

Bien que tous les outils Google pour les webmasters ne soient pas adaptés à la recherche en entreprise et au système GSA, il est utile de suivre les consignes offertes dans l'[aide de ces outils](#), car elles permettent d'appréhender la manière dont le contenu est classé par un algorithme de recherche Google.

Création et structure du contenu

Le système GSA est compatible et indexe de nombreux types de documents. Vous devez donc passer en revue différentes consignes pour préparer au mieux l'indexation d'un type de document donné.

Cette section fournit les instructions à suivre pour préparer l'indexation des types de documents suivants :

- [Documents Microsoft Office](#)
- [Documents PDF](#)
- [Documents HTML](#)

Des consignes concernant les [flux de documents](#) et [la classification/taxonomie des données](#) sont également incluses.

Elles traitent principalement de l'indexation de métadonnées supplémentaires avec le contenu des documents. Les métadonnées peuvent enrichir le contenu figurant dans l'index du système GSA. Pour obtenir l'exemple d'un type d'attribut de métadonnées pouvant être ajouté pour enrichir la classification des documents, référez-vous au schéma de métadonnées [Dublin Core](#) . Ce dernier propose un ensemble de champs textuels permettant de décrire une ressource.

Les propriétaires de contenus peuvent suivre certains principes lors de la création de documents qui permettront de distinguer efficacement et d'identifier de manière unique leurs documents. Ce point est important, car il manque aux documents d'entreprise une structure d'URL définie , pourtant utile pour identifier les relations sur la toile. C'est la raison pour laquelle il est important d'appliquer un identifiant unique aux documents figurant dans un corpus de documents d'entreprise.

Documents Microsoft Office

Dans les propriétés d'un document Microsoft Office, renseignez les champs suivants :

- Auteur
- Titre
- Commentaires

Si de nombreux documents sont créés à partir de modèles, il est très important de changer les propriétés par défaut du document.

Documents PDF

Dans les propriétés d'un document PDF, renseignez les champs suivants :

- Auteur
- Titre
- Créateur

Si le document n'est pas directement autorisé par un outil de publication de PDF, assurez-vous que ces propriétés sont renseignées dans le document source avant de convertir ce dernier en PDF.

Documents HTML

Dans un document HTML, incluez les informations suivantes :

- <title> description
- du contenu (saisi manuellement ou généré automatiquement) : assurez-vous que les descriptions sont de qualité, même si elles ne s'affichent pas sur la page elle-même.

Veillez à ce que les descriptions de pages distinctes soient différentes. Pour limiter l'indexation du contenu d'une page donnée, utilisez la balise googleoff.

Flux de documents

Le protocole de flux GSA permet l'indexation de métadonnées avec le contenu. Incluez dans le flux de document toute autre métadonnée qui n'existe pas ailleurs dans le contenu et qui peut enrichir le contenu indexé.

Classification/Taxonomie des données

Si des documents devant être indexés par le système GSA sont associés à un schéma de classification et/ou une taxonomie des données d'entreprise, assurez-vous que les informations d'ontologie sont bien indexées par le système GSA. La transmission des informations d'ontologie avec le contenu n'est pas obligatoire, mais peut améliorer la solution de recherche.

La taxonomie peut être utilisée en association avec les filtres de métadonnées pour restreindre les recherches sur les documents à certains niveaux de la hiérarchie globale.

Selon le type de document et sa source, l'indexation des informations de taxonomie peut demander un processus d'intégration associant la classification du document et le document en lui-même dans l'index. Une recherche poussée est nécessaire pour savoir comment associer ces informations au système GSA au moment de l'indexation.

Reconnaissance d'entités

La fonctionnalité de reconnaissance d'entités du système GSA peut être utilisée pour ajouter automatiquement des métadonnées aux documents au moment de l'indexation. Cet ajout s'effectue selon des règles prédéfinies contenues dans des dictionnaires de textes ou via une structure XML qui effectue des comparaisons en fonction de termes ou d'expressions régulières. Les termes professionnels courants ou standards au sein de votre organisation sont particulièrement adaptés à la définition d'entités. Si une taxonomie d'entreprise a été identifiée, il est également intéressant de l'inclure dans les règles de reconnaissance d'entités. Ainsi, même si certains documents ne sont pas identifiés avec ces termes, GSA peut effectuer l'association au moment de l'indexation grâce à ces règles.

Options de réglage de la pertinence

Google Search Appliance comporte des fonctionnalités qui permettent à l'administrateur d'un système de recherche de modifier légèrement la pondération des résultats pour toute l'organisation ou simplement pour quelques utilisateurs. Le classement des résultats de recherche peut être pondéré en fonction d'une date, d'une métadonnée, d'une source, de collections ou d'une source de flux.

Par exemple, imaginons que plusieurs sources de contenu alimentent le système GSA et que SharePoint en fasse partie. Vous avez remarqué que, contrairement à ce qui est souhaité, les résultats SharePoint ne sont pas suffisamment bien placés dans la liste des résultats. Vous pouvez donc décider de pondérer la source de contenu SharePoint afin d'améliorer le classement de son contenu.

Pour en savoir plus sur la manière de configurer une règle de pondération de résultats, consultez la page d'aide [Traitement > Pondération des résultats](#).

Le tableau suivant présente les fonctionnalités qui peuvent avoir des conséquences sur la manière dont les résultats sont présentés aux utilisateurs.

| Fonctionnalité | Commentaires |
|----------------------------------|---|
| Pondération de la source | Avec la correspondance de format, pondérez une source par rapport à l'autre. |
| Pondération de la date | Attribuez plus ou moins d'importance à la date de création des documents. |
| Pondération des métadonnées | Pondérez les documents associés à des métadonnées spécifiques. |
| Correspondances | Bien que les correspondances soient une fonctionnalité de suggestion et ne fassent pas partie des résultats de recherche naturelle, vous pouvez les utiliser pour placer des documents en tête de liste pour certaines requêtes. |
| Extension des requêtes | Utilisez une règle d'extension des requêtes pour élargir les termes de recherche à d'autres termes (synonymes). |
| Marqueur intelligent | Lorsque les rapports détaillés sur les recherches sont activés, le système GSA utilise la fonctionnalité "Marqueur intelligent" pour analyser les données de flux de clics et placer certains résultats de recherche en tête au fur et à mesure. Par exemple, pour une requête de recherche donnée, si les utilisateurs cliquent tout le temps sur le second résultat au lieu du premier, c'est ce second résultat qui finira par prendre la première position sur la page. |
| Regroupement et filtrage d'hôtes | Autorisez le système GSA à filtrer toute combinaison de : <ul style="list-style-type: none"> • résultats provenant du même chemin ; • résultats comportant le même titre et le même extrait. |
| Structure de classement | Spécifiez une pondération par URL. Veuillez noter qu'il s'agit d'une solution très complexe à gérer. Elle doit toujours être tentée en dernier recours. |
| Mots vides | Utilisez les mots vides pour supprimer certains termes de la requête afin qu'ils ne soient pas utilisés dans la recherche. Veuillez utiliser cette fonctionnalité avec précaution, car elle peut avoir d'importantes répercussions sur le classement si elle permet de résoudre un problème spécifique. |
| Collections | Répartissez le contenu dans plusieurs collections afin de limiter le corpus de documents pour une requête de recherche. |

Points à prendre en compte concernant l'expérience utilisateur

Les modifications apportées à l'expérience utilisateur en matière de recherche peuvent également modifier la manière dont vos utilisateurs perçoivent le déploiement du système de recherche.

Pour obtenir des informations sur les bonnes pratiques relatives aux options permettant d'améliorer l'expérience de recherche des utilisateurs, consultez la section [Bonnes pratiques en matière d'expérience de recherche](#).

Pour obtenir des informations détaillées sur les options relatives à l'expérience utilisateur, consultez [GSA en pratique : Guide d'expérience utilisateur](#).

Chapitre 2 : Appartenance du système GSA et modèle d'imputation des coûts

Présentation

Trois modèles principaux d'appartenance d'une solution de recherche s'appliquent aux déploiements GSA :

- [Appartenance à une organisation informatique centralisée](#)
- [Appartenance à une seule unité fonctionnelle ou commerciale](#)
- [Appartenance à un tiers dans un environnement hébergé](#)

Chaque modèle dispose de caractéristiques spécifiques et a ses avantages et ses inconvénients selon l'environnement de votre organisation. Pour choisir un modèle qui répond à vos besoins, prenez en compte les points suivants :

- Modèle de financement et d'imputation des frais
- Définition du processus
- Responsabilité
- Données utilisées pour établir les rapports sur l'utilisation et/ou sur la valeur
- Modèle de création de rapports
- Contrats de fournisseurs tiers existants pour l'hébergement et/ou les services

Appartenance à une organisation informatique centralisée

Dans les organisations dont le modèle informatique est centralisé, il est fréquent que la solution de recherche GSA soit rattachée à un centre de services partagés informatiques. Dans ce modèle, un groupe de ressources principal rassemble les connaissances relatives aux fonctionnalités du système GSA et aux impacts commerciaux de la mise en œuvre des fonctionnalités et de l'intégration des sources de contenu.

Ce groupe connaît également le champ d'application du contexte de l'entreprise dans son ensemble, et non d'une unité fonctionnelle ou commerciale donnée. Le groupe assume les fonctions suivantes :

- Identification et planification d'une approche de déploiement par phases du système de recherche dans l'organisation.
- Prise en compte des requêtes soumises par les unités commerciales et les propriétaires de contenus qui souhaitent que ces derniers soient indexés dans le système GSA et accessibles par les utilisateurs de toute l'entreprise.

L'appartenance à un seul groupe au sein d'une organisation permet également de simplifier la gestion des comptes et l'accès au système de recherche.

Modèles de coûts et d'imputation des frais

Il existe de nombreuses façons de calculer comment financer et soutenir la solution de recherche :

- Fournir les fonds nécessaires au centre de services partagés dédié à la recherche, sans facturer les unités commerciales qui utilisent le système GSA.
- Imputer les frais de déploiement et de maintenance du système GSA aux unités commerciales qui l'utilisent :
 1. Calculez le coût de base de GSA (frais de licences, centre de données, installation, maintenance, etc.).
 2. Évaluez la part du système GSA utilisée par une unité commerciale afin de trouver une valeur à facturer en multipliant ce pourcentage par le coût total du système GSA. L'utilisation peut être estimée en fonction du pourcentage de licences de documents utilisées ou de l'utilisation des requêtes de recherche et/ou de leur nombre.
 3. Facturez la configuration/personnalisation du système GSA sur une base horaire et matérielle (chaque heure travaillée doit être comptabilisée).

Pour la personnalisation, notamment pour les modifications SXLT des frontaux, un modèle hybride semble être le plus efficace, car l'unité commerciale peut avoir une idée précise du type de système de recherche attendu.

Appartenance à une seule unité fonctionnelle ou commerciale

Dans les organisations dont le modèle informatique est décentralisé, il est fréquent que des unités fonctionnelles ou commerciales individuelles gèrent la solution de recherche GSA en fonction de leurs besoins spécifiques. Les unités commerciales connaissent bien leur système, leurs processus et les formats d'utilisation des utilisateurs. Elles sont donc très bien placées pour déployer une solution de recherche qui soit la plus utile possible pour leurs utilisateurs. À un autre niveau, demander aux utilisateurs finaux de donner leur avis sur la solution peut être précieux, car leurs besoins sont alors mis au centre des préoccupations.

Placer la solution de recherche entre les mains d'unités commerciales individuelles peut également permettre de remplacer les processus de gouvernance informatique habituels et d'accélérer le délai de mise sur le marché. Dans le meilleur des cas, le déploiement du système GSA par une unité commerciale peut servir de modèle pour étendre l'utilisation du système GSA à l'ensemble de l'entreprise.

Même si les unités fonctionnelles ou commerciales peuvent ne pas suivre les processus informatiques standards, il est important de se concentrer sur les points suivants et de les passer en revue afin d'évaluer le succès du déploiement de la solution de recherche :

- Financement
- Définition de quelques processus basiques
- Mettre en place un modèle d'assistance à l'avenir
- Données à utiliser pour établir les rapports d'utilisation et/ou de valeur
- Modèle de création de rapports

Même si le système de recherche peut ne pas être déployé selon les pratiques informatiques standards, il est très important d'avoir des données et une ligne de conduite qui peut servir de base à de futures améliorations du système de recherche.

En fonction de l'environnement informatique de votre organisation, ce modèle peut avoir pour conséquence la non-conformité du déploiement du système GSA avec certaines règles de gouvernance informatique en vigueur, dans des domaines tels que l'audit, le branding, la plate-forme, l'architecture, etc. Un système de recherche déployé au sein d'un tel modèle peut être implicitement limité et ne pas pouvoir s'étendre aux référentiels contrôlés hors de l'unité commerciale. Ce modèle s'adapte particulièrement à des applications de recherche développées dans un but bien spécifique et destinées à un groupe d'utilisateurs donné (qui peut ne pas être représentatif de l'entreprise dans son ensemble).

Modèle de coûts

Le déploiement du système GSA par une unité commerciale individuelle peut être financé en calculant le coût d'acquisition du système GSA, ainsi que les ressources nécessaires pour créer la solution. Un modèle d'assistance doit également être développé.

Même si la solution de recherche est destinée à être déployée et détenue par une unité commerciale, il peut être plus efficace d'étudier certains processus et d'externaliser l'assistance continue à un tiers, selon les pratiques de l'entreprise en matière d'externalisation (en interne ou par un tiers).

Appartenance à un tiers dans un environnement hébergé

Il peut être justifié pour une organisation d'externaliser la solution de recherche GSA à un fournisseur d'hébergement tiers. Ce modèle peut être défini par les deux approches générales suivantes :

- Définir et développer la solution en interne et déléguer la partie opérationnelle à un tiers.
- Obtenir les services d'un tiers pour permettre de définir, de développer et d'héberger la solution ou choisir une offre d'hébergement de la solution de recherche GSA complète par un tiers.

Un tel contrat avec un tiers doit être conclu selon les consignes de recrutement des fournisseurs, actuelles ou passées.

L'avantage de ce modèle réside dans le fait qu'un fournisseur tiers spécialisé dans la recherche GSA dispose déjà des ressources et de l'infrastructure permettant de développer et de fournir la solution de recherche. Cette approche a également l'avantage de réduire les coûts. Le manque de flexibilité et la difficulté d'intégration dans les systèmes existants sont les deux inconvénients. Les environnements hautement sécurisés ayant des exigences relatives au stockage sur site de données sensibles semblent ne pas convenir à ce type d'approche.

Modèle de coûts

Vous devez étudier les modèles de coûts d'appartenance à des unités commerciales spécifiques et d'appartenance à une organisation informatique centralisée et les comparer au coût proposé pour une solution d'hébergement par un tiers. Ce modèle est régi par les conditions d'utilisation du contrat d'un fournisseur. Tous les aspects doivent donc être négociés et définis en amont.

Le modèle d'assistance, ainsi que les contrats de niveau de service applicables sont des éléments importants à faire figurer au contrat. En fonction de ce dernier, ce modèle peut être plus intéressant en termes de coûts en amont, mais offre moins de flexibilité quant à l'évolution de la solution.

Gestion des opérations et assistance continues

Comme décrit brièvement dans les sections précédentes, la solution de recherche GSA doit faire l'objet d'une maintenance de façon à être évolutive.

En premier lieu, l'aide proposée par l'assistance Google for Work est disponible en cas de problème pour toute la durée de votre licence GSA. En outre, une personne ou un groupe familiarisé avec le système de recherche au sein de votre organisation doit être le référent. Cette personne ou ce groupe doit :

- connaître l'architecture de recherche du système GSA telle que déployée au sein de votre organisation ;
- connaître les fonctionnalités GSA et la manière dont elles sont configurées dans la fonctionnalité d'administration du système GSA (via la Console d'administration ou par l'intermédiaire des API Admin) ;
- être le gardien de tout processus ou de toute norme mis en place pour embarquer le nouveau contenu dans le système GSA ;
- coordonner et planifier les éventuelles améliorations proposées relatives aux frontaux et à l'expérience utilisateur ;
- exécuter des rapports de recherche GSA ou fournir une assistance à toute personne demandant un rapport de recherche GSA ;
- rester informé de toute nouvelle version de GSA et planifier les mises à jour régulières du système (une fois par an, par exemple) ;
- être le référent en cas de problème rencontré avec GSA et déposer des demandes d'assistance auprès de Google si nécessaire.

Pilotage continu de la recherche

Quel que soit le type de modèle d'appartenance de la solution de recherche, il est important que les principaux membres de l'équipe responsable se réunissent régulièrement pour en évaluer l'efficacité et définir les actions prioritaires. Par exemple, une équipe composée des commanditaires du système GSA, des propriétaires du système, des administrateurs et des utilisateurs finaux partie prenante peut se réunir tous les trimestres (ou plus fréquemment si la solution de recherche est un domaine hautement prioritaire pour l'entreprise) afin de passer en revue les commentaires des utilisateurs et les statistiques de recherche, et d'établir les priorités concernant les activités d'amélioration ou les domaines destinés à être intégrés dans la solution de recherche.

Les analyses et les statistiques, ainsi que la manière dont elles influencent le perfectionnement de la solution représentent d'autres aspects importants dans le pilotage continu de la solution de recherche. La solution de recherche peut être un important indicateur commercial, car elle peut permettre de rassembler des statistiques et des données.

L'intégration dans un service d'analyse tel que Google Analytics ou dans une solution personnalisée peut fournir des statistiques telles que : impressions de l'utilisateur, points faibles, sujets fréquemment recherchés, indications utiles au développement de produits, etc. Le fait d'avoir une stratégie de collecte et d'analyse des données de recherche permet de tirer davantage parti de la solution de recherche dans des domaines insoupçonnés au premier abord.

Chapitre 3 : Modèles d'acquisition de contenu de GSA

Présentation

Ce chapitre présente des stratégies et des approches permettant d'intégrer les sources de contenu de votre entreprise au système GSA.

Bien que le système GSA, dans de nombreux cas, soit capable de découvrir, d'explorer et d'indexer du contenu sans difficulté, il est préférable de planifier l'intégration et de mettre au point une stratégie sur la manière d'intégrer différentes sources de contenu et/ou applications. Des éléments tels que la sécurité, le type d'intégration, les métadonnées disponibles, la définition des entités pour la reconnaissance d'entités et l'impact sur l'interface utilisateur jouent un rôle important dans le succès de l'intégration.

Prise en compte des sources de contenu pour l'indexation

Au moment de choisir les plates-formes à intégrer au système GSA, réfléchissez aux points figurant dans le tableau suivant et attribuez des priorités aux différentes plates-formes en conséquence.

| Éléments de réflexion | Commentaires |
|---|---|
| Utilisateurs et types d'utilisateurs concernés par l'intégration du contenu | Optez en priorité pour les plates-formes qui ont le plus d'impact. |
| Simplicité d'intégration | Par ordre de difficulté croissante d'intégration : exploration directe, connecteur disponible, exploration par l'intermédiaire d'un proxy, nécessité de développer des flux, nécessité de développer des connecteurs. |
| Mécanisme de sécurité d'autorisation et d'authentification requis | <p>Les systèmes s'appuyant sur des mécanismes de sécurité directement compatibles avec le système GSA sont les plus faciles à intégrer.</p> <p>Les plus difficiles à intégrer sont ceux qui requièrent l'intégration d'un mécanisme d'authentification personnalisé.</p> |
| Personnalisation des frontaux de recherche requise pour que l'expérience de recherche de la plate-forme ait le plus de valeur | <p>Il arrive que, pour tirer au maximum parti de l'intégration du contenu au système GSA, il soit nécessaire de modifier les frontaux afin de mettre à disposition des éléments tels que les suivants :</p> <ul style="list-style-type: none"> ● Filtres ● Catégories ● Métadonnées dans les résultats de recherche ● Processus de navigation personnalisés <p>Par ordre de difficulté croissante d'intégration : XSLT par défaut, XSLT avec petites modifications, XSLT avec modifications importantes, analyse personnalisée d'application et affichage du XML fourni par le système GSA.</p> |
| Maturité des métadonnées disponibles dans la source de contenu | <p>Les métadonnées ne sont pas requises pour l'indexation. En revanche, elles peuvent permettre d'enrichir le contenu figurant dans l'index et offrir d'autres options de mise en forme de l'interface de recherche.</p> <p>Certaines sources de contenu, en raison de leur nature, disposent de métadonnées "prêtes à l'emploi", alors que d'autres nécessitent l'adhésion à un processus au moment de la publication.</p> <p>Le cas échéant, envisagez également d'ajouter des métadonnées aux documents, par programmation, au moment de l'indexation, via un processus personnalisé.</p> |

| | |
|--|--|
| Ajouter des métadonnées aux documents par l'intermédiaire de la reconnaissance d'entités | La fonctionnalité de reconnaissance d'entités du système GSA peut être utilisée pour enrichir les contenus par l'extraction d'entités en définissant des règles d'entité basées sur des dictionnaires ou des expressions régulières. Ces règles, lorsqu'elles sont définies, associent des métadonnées aux documents au moment de l'indexation. Cette fonctionnalité peut être utilisée pour attribuer des métadonnées aux documents qui, à défaut, ne seraient associés à aucune métadonnée. |
|--|--|

Référentiel/plate-forme de contenu et modèle d'acquisition d'application d'entreprise

Deux approches de haut niveau doivent être prises en compte pour la sélection du contenu à intégrer au système GSA :

- [Intégration du contenu par plate-forme](#)
- [Intégration du contenu par application d'entreprise](#)

Différents points doivent être soumis à réflexion en fonction de ces modèles. Certains de ces points sont décrits dans les sections qui suivent.

Intégration du contenu par plate-form

Avec l'intégration de contenu par plate-forme, tous les documents d'un référentiel de contenu (ou la plupart) doivent être indexés sans tenir compte des applications concernées par ce référentiel. Le modèle d'intégration de contenu par plate-forme est adapté à une organisation qui centralise le financement de la solution de recherche et lorsque cette dernière est rattachée à une équipe de recherche centralisée.

Dans ce modèle, le contenu est intégré à un référentiel (ou plate-forme) après l'autre, sans tenir compte des applications ou des unités commerciales détenant du contenu dans la plate-forme. Cette approche facilite également le déploiement du concept de "recherche universelle" dans l'organisation, car la solution est considérée de manière holistique et non application par application.

L'indexation de contenu issu de Livelink ECM en est un exemple. Dans cette approche, la plate-forme Livelink est indexée en totalité sur le système GSA au cours d'un seul projet de déploiement, quelles que soient les unités commerciales ou applications qui stockent les documents dans le système.

Vous devez attribuer des priorités aux sources de contenu que vous souhaitez intégrer en tenant compte des facteurs répertoriés dans la section [Prise en compte des sources de contenu pour l'indexation](#).

Intégration du contenu par application d'entreprise

Avec l'intégration de contenu par application d'entreprise, tous les documents (ou la plupart) d'une application métier spécifique sont indexés, sans tenir compte de(s) plate-forme(s) à la source du contenu. La mise en place d'un déploiement GSA en choisissant des applications métier spécifiques pour l'intégration est adapté à une organisation où la recherche est gérée de manière centralisée, mais financée par les unités commerciales concernées.

Ce modèle convient également lorsqu'une unité fonctionnelle d'entreprise (et non une équipe de services partagés) est responsable de la mise en œuvre de la solution de recherche GSA. Dans cette approche, les applications métier spécifiques sont ciblées et indexées en priorité sur le système GSA.

L'indexation du contenu d'un système de gestion du recrutement RH est un bon exemple de cette approche, lorsque les applications du système sont stockées dans un système de gestion de contenu d'entreprise. Le choix de l'intégration par application plutôt que par plate-forme permet d'indexer uniquement ces applications de recrutement et non le système de gestion de contenu dans son ensemble.

Les applications à indexer doivent se voir attribuer des priorités, comme pour les plates-formes, à l'aide des facteurs répertoriés dans la section [Prise en compte des sources de contenu pour l'indexation](#).

Un des principaux objectifs de l'indexation de contenu d'applications dans le système GSA est d'accroître la productivité des utilisateurs de cette application. Lors de la réflexion sur des cas d'utilisation ou sur le flux de travail des utilisateurs finaux, identifiez les domaines dans lesquels le système GSA pourrait simplifier le processus.

Par exemple, dans une application où les documents peuvent être modifiés, les résultats de recherche peuvent afficher le document, mais aussi un lien vers le mode d'édition du document pour que les utilisateurs qui souhaitent en modifier le contenu puissent accéder directement à ce mode sans passer par une étape intermédiaire. Toute optimisation ou simplification des processus utilisateur par le système GSA est bénéfique.

Processus continu d'acquisition et de maintenance

La solution de recherche GSA continue à évoluer en même temps que l'activité, au fur et à mesure que d'autres contenus sont créés au sein de l'entreprise. Une organisation doit se préparer à gérer de futures requêtes sur le contenu intégré sur le système GSA au-delà des phases de déploiement initiales. Vous apprendrez beaucoup du premier déploiement et ce dernier continuera à vous servir pour effectuer des ajouts ou des ajustements et faire évoluer la solution.

Il n'est pas nécessaire d'établir un processus élaboré, mais nous conseillons de garder à disposition un modèle de spécification de recherche que vous pourrez utiliser pour vous guider dans la conception et la mise en œuvre de l'intégration du contenu. Ce document de spécification doit contenir des informations techniques et fonctionnelles sur l'intégration. Sans document pour vous guider, le champ d'application de la mise en œuvre peut devenir ingérable et les exigences peuvent être perdues et jamais respectées.

Google recommande de suivre les étapes de processus décrites dans les sections suivantes :

1. [Analyse initiale du contenu](#)
2. [Développement/Configuration](#)
3. [Test et déploiement](#)

Veillez noter que le processus recommandé ne prend pas en compte les tâches d'allocation de ressources et de gestion du projet nécessaires dans une telle situation.

Analyse initiale du contenu

L'investigation initiale pour l'intégration de contenu au système GSA est déclenchée par une demande directe ou fait suite à l'identification d'un domaine dans lequel le contenu n'est pas intégré au système GSA et n'est pas renvoyé dans les résultats de recherche.

Lors de cette étape, effectuez l'analyse suivante :

1. Analysez les facteurs d'intégration présentés dans la section [Prise en compte des sources de contenu pour l'indexation](#).
2. Répertoriez les exigences et remplissez la partie "spécification de recherche" pour le système concerné.
 - a. Étudiez les exigences de mise en forme de l'interface utilisateur.
 - b. Veillez à déterminer comment le contenu nouvellement acquis et les spécifications relatives à l'interface utilisateur vont être intégrés à l'interface et aux contenus existants.
3. Analysez les métadonnées disponibles dans le système identifié et la manière dont elles peuvent élaborer et améliorer la solution de recherche.
 - a. Prenez en compte toutes les actions requises pour supprimer les métadonnées ou ajouter les métadonnées au contenu, le cas échéant.
 - b. Définissez des entités pour la reconnaissance d'entités afin d'ajouter les métadonnées qui ne seraient pas disponibles sans cette manipulation.
4. Identifiez les contacts référents qui agiront comme de petites entreprises à l'avenir.

Développement/Configuration

Au cours de la phase de développement/configuration, suivez les étapes suivantes :

1. Intégrer le contenu : cette étape inclut le déploiement de tout composant nécessaire extérieur au système GSA tel que les connecteurs, SAML Bridge, etc.
2. Développer les composants d'interface utilisateur requis.

Test et déploiement

Au cours de la phase de test et de déploiement, suivez les étapes suivantes :

1. Conduire un test d'intégration/du système : le contenu est-il indexé et affiché correctement ?
2. Conduire un test de pertinence : comment le contenu nouvellement acquis s'intègre-t-il au contenu existant dans la collection "par défaut" ou "universelle" regroupant le contenu issu de nombreuses sources ?
3. Conduire un test ou une analyse de charge/performances : comment la charge supplémentaire va-t-elle affecter les performances de la solution existante ?

Gestion de données sensibles nouvellement indexées dans le système GSA

GSA est un outil de recherche puissant, qui permet aux utilisateurs de trouver des documents qu'ils n'avaient auparavant aucun moyen de trouver. Dans certains cas, juste après le déploiement, les utilisateurs ont commencé à trouver, dans leurs résultats de recherche, des documents auxquels ils n'auraient pas dû avoir accès. Ces situations se sont produites alors que le système GSA respecte bien les autorisations sur les documents, en raison d'autorisations du système source trop permissives. L'autorisation de ces documents doit être restreinte pour que ces derniers ne s'affichent pas dans les résultats de recherche.

Lors du déploiement du système GSA, une bonne pratique consiste à appliquer un plan de mitigation prévoyant la fuite de données sensibles. Cette opération permet de planifier et d'atténuer les impacts par le développement de processus, la communication et la formation des utilisateurs au sein de votre organisation.

Le contenu RH est souvent celui le plus exposé à ce type de risques. Envisagez un simple audit d'un échantillon de données RH pour vérifier les autorisations actuelles. Envisagez de ne pas inclure les contenus sensibles lors de la première phase de déploiement afin de laisser le temps aux utilisateurs et aux administrateurs de se familiariser avec le système GSA. Les administrateurs seront ainsi mieux préparés pour traiter d'éventuelles fuites de données non intentionnelles dues à l'absence de restrictions des autorisations sur les documents.

L'exemple de processus suivant permet de traiter les incidents qui surviennent lorsque le système GSA semble renvoyer des résultats de recherche sensibles dont les autorisations auraient dû être restreintes :

- Configurez un alias d'adresse e-mail permettant de gérer la création de rapports utilisateur lorsque l'accès aux documents devrait être restreint.
- Désignez un rôle d'analyse d'affaires en tant que ressource référente pour la gestion de chaque cas qui se présente.
 - Cette personne doit pouvoir relier le document en question à la source de contenu à laquelle il appartient.
 - Elle doit également pouvoir communiquer efficacement avec les propriétaires de la source de contenu afin de déterminer le paramètre d'autorisation le plus approprié pour le document concerné.

- Avant de mettre le système à disposition, le processus de gestion des documents devant être restreints doit faire l'objet d'une communication claire.
 - Les utilisateurs doivent recevoir l'alias d'adresse e-mail à laquelle envoyer les incidents rencontrés.
 - Les utilisateurs doivent recevoir un exemple d'informations devant être fournies lors de la transmission d'un incident.

- L'analyste d'affaires doit déterminer la sévérité de tous les rapports entrants et avoir la possibilité de supprimer immédiatement le document, de manière permanente ou temporaire, des frontaux du système GSA. Il peut s'agir d'une suppression temporaire jusqu'à ce que les propriétaires du contenu aient confirmé les niveaux d'accès.

- Le suivi des incidents peut être effectué dans le système de gestion des demandes de votre choix.

Chapitre 4 : Approche de l'environnement de GSA

Présentation

Il est intéressant de tester les changements dans un environnement distinct avant de les déployer pour les utilisateurs finaux. Comme pour tout type de serveur ou d'application, un changement si minime soit-il dans la configuration peut avoir des conséquences inattendues. Une stratégie de test en bonne et due forme et un environnement de transition sont donc recommandés en cas de modifications apportées à n'importe quelle application de recherche.

Les environnements suivants sont recommandés pour réussir le développement, le test, la mise en œuvre et le déploiement en production de la solution :

- [Environnement de développement/Fonctionnalités expérimentales](#)
- [Environnement de test/assurance qualité](#) (facultatif)
- [Environnement de production](#)

Bien que ce soit la solution idéale, les environnements physiques ne sont pas indispensables. Par exemple, les fonctionnalités expérimentales/de développement et le test/assurance qualité peuvent être déployés sur le même système GSA en utilisant différents frontaux et collections.

Environnement de développement/fonctionnalités expérimentales

Pour Google Search Appliance, un environnement de développement implique simplement de répliquer l'environnement de production afin de fournir une plate-forme distincte pour développer et intégrer les modifications apportées à la configuration et les nouvelles améliorations. L'environnement de développement doit inclure l'accès aux mêmes types et sources de contenu que l'environnement de production, mais doit comporter un ensemble limité ou réduit de documents. Tous les composants d'intégration déployés (SAML Bridge ou connecteurs, par exemple), doivent avoir des instances correspondantes dans l'environnement de développement.

L'environnement de développement n'a pas besoin d'avoir la taille requise par les normes de production. Il peut être réduit en conséquence, car il n'a pas besoin de prendre en compte les exigences relatives à la taille des documents ou aux performances d'un système de production.

Une configuration de développement courante inclut un système de recherche non déployé en production qui n'affiche pas les résultats à la plupart des utilisateurs finaux. Toutes les modifications, mises à jour et améliorations apportées à la configuration sont testées sur ce système de recherche, puis transmises à la prochaine étape d'environnement dès qu'elles sont validées.

Si possible, évitez d'utiliser des conventions de nommage codées en dur qui peuvent compliquer la migration des configurations. Par exemple, `dev_collection` ou `test_frontend` devront être renommées pour passer en production.

Les API Google Search Appliance Admin, présentes à partir de la version 6.0 du logiciel, vous permettent d'exporter et d'importer des configurations détaillées (correspondances individuelles, par exemple) ou générales (configuration complète d'un système de recherche). Pour en savoir plus sur les API Admin, consultez les Guides des API d'administration pour les développeurs, disposant de liens vers la [page de documentation GSA publique](#).

L'environnement de développement peut également être utilisé en tant qu'environnement de test avant le lancement en production des exigences réalisées.

Environnement de test et d'assurance qualité (facultatif)

L'environnement de test et d'assurance qualité est similaire à celui de développement, puisqu'il s'agit d'un environnement répliqué qui n'est pas en production. Cet environnement doit rester stable pour toutes les personnes effectuant des tests sur le système. L'ajout de composants de déploiement, les modifications apportées à la configuration ou aux frontaux doivent être effectués par étapes, conformément à un processus de gestion de la configuration. L'équipe responsable des tests doit être informée de tout changement intervenant dans cet environnement qui pourrait affecter leurs tests.

Le système GSA est un produit qui indexe les sources de contenu sans les modifier, l'environnement d'assurance qualité GSA peut donc pointer vers des sources de contenu dans l'environnement d'assurance qualité ou dans l'environnement de production, avec différents points à prendre en considération pour les deux approches.

Les tests de performance peuvent être exécutés dans cet environnement afin de développer une trame, en prenant en compte le fait que l'environnement de production le dépassera en termes de spécifications matérielles.

Environnement de production Production

L'environnement de production est celui utilisé par les utilisateurs. Sa taille doit être adaptée aux exigences relatives à la taille de l'index et aux performances.

Les modifications apportées à l'environnement de production doivent être introduites par l'intermédiaire d'un processus strict de gestion de la configuration. Toutes les modifications transférées en production doivent avoir été testées et validées dans l'environnement d'assurance qualité. Les interruptions opérationnelles doivent être prises en compte lors de la migration en production. Les modifications doivent être effectuées au moment où elles auront le moins d'impact sur l'activité de l'entreprise.

Environnement de sauvegarde de la production/récupération en cas de sinistre

Dans la plupart des cas, notamment quand la solution de recherche est une application critique pour les utilisateurs, il est important d'avoir un environnement de sauvegarde/récupération en cas de sinistre dans l'environnement de production. Un environnement de sauvegarde/récupération en cas de sinistre est une copie de l'environnement de production et peut être utilisé en cas d'interruption, planifiée ou non. Cet environnement est en principe une copie exacte de l'environnement de production et dispose de la même infrastructure et de la même capacité.

Points à prendre en compte pour la planification de la capacité de l'environnement de production.

Les environnements de production répliqués ne sont pas toujours uniquement utilisés pour leurs fonctions de sauvegarde ou de récupération suite à un sinistre. Ils peuvent également servir à la planification de la capacité. Si un système GSA ne peut pas gérer toute la charge de recherche au moment de la présentation des résultats, la réplique d'un GSA peut être utilisée dans une configuration de charge répartie dans le but de planifier la capacité. Ainsi, la charge de traitement est distribuée entre plusieurs GSA traitant le même contenu.

Veillez noter qu'en raison de l'existence de la fonctionnalité de mise en miroir du système GSA, un seul système GSA récupère les nouveaux contenus, puis transmet les modifications aux systèmes répliqués et les synchronise.

Récapitulatif

La gouvernance n'est pas un sujet universel et diffère selon l'organisation. Les sections de ce document présentent quelques concepts que vous pouvez adopter lors de la définition du type de gouvernance de votre solution de recherche GSA nouvellement déployée. Gardez à l'esprit qu'il existe de multiples façons d'intégrer ces concepts à vos processus existants.

Comme toujours et comme pour l'architecture autour du système GSA, essayez de concevoir des processus de gouvernance aussi simples que possible pour que les déploiements soient rapides et itératifs. Essayez d'utiliser le système GSA pour vous servir de processus informatiques existants qui vous permettront de déployer rapidement et de reproduire la conception de la solution de recherche sur le système GSA.