SPOTLAB / ANALYSE P14

RAG vs GraphRAG : quelle architecture choisir pour les connaissances d'entreprise ?

Une matrice pour choisir l'architecture de recherche documentaire adaptée à la nature des questions et du corpus.

Analyse / P14

Novembre 2026

Sources consultées au 2 novembre 2026

ANALYSE P14

Sommaire

01	Trois enseignements pour décider	3
02	RAG et GraphRAG ne répondent pas exactement aux mêmes questions	4
03	Ce que démontrent — et ne démontrent pas — les recherches	6
04	La méthode Spotwork.AI : cinq décisions avant d'investir dans un graphe	7
05	Gouvernance des connaissances : les points souvent oubliés	8
06	Exemple : analyser un référentiel de conformité	9
07	Ce qu'un décideur doit retenir	10
08	Sources et limites	10

FICHE DE L'ANALYSE

ÉDITEUR
Spotwork.AI — Spotlab, Research & Insights

RÉFÉRENCE
P14

TYPE
Analyse documentaire

COLLECTION
Données & connaissances

SOURCES CONSULTÉES
2 novembre 2026

PUBLICATIONS LIÉES
I0-03 · P15 · P16

EN BREF

Une entreprise peut disposer de milliers de documents et pourtant peiner à obtenir une réponse fiable. Faut-il améliorer son moteur RAG, construire un graphe de connaissances ou combiner les deux ? Le choix dépend d'abord des questions à traiter et des preuves attendues, pas de la sophistication de l'architecture.

> *Les approches RAG et GraphRAG doivent être évaluées sur les questions réelles et la qualité du corpus, plutôt que sur leur seule sophistication technique.*
>
> CE QU'IL FAUT RETENIR

01 Trois enseignements pour décider

1

Pour retrouver une information précise, un RAG bien configuré reste un point de départ sérieux.

La segmentation, la recherche hybride, le reclassement et les permissions comptent parfois davantage que l'ajout d'un graphe.

2

Pour comprendre des relations ou des tendances à l'échelle d'un corpus, GraphRAG mérite une évaluation.

Les graphes d'entités et les synthèses de communautés ouvrent d'autres chemins de recherche, sans garantir une meilleure réponse pour toutes les tâches.

3

La qualité et le coût se jugent sur l'ensemble du système.

Construction, mises à jour, provenance des liens, fidélité des réponses, contrôle des accès et effort de revue doivent entrer dans la décision.

DÉCISION RECOMMANDÉE

Mener un pilote comparatif par famille de questions, sur un corpus autorisé, avec les mêmes sources et un protocole commun. **N'adopter la couche graphe que lorsque ses bénéfices sont démontrés pour des tâches suffisamment importantes pour justifier son coût opérationnel.**

02 RAG et GraphRAG ne répondent pas exactement aux mêmes questions

Le RAG (Retrieval-Augmented Generation) recherche des passages dans une base documentaire avant de les transmettre au modèle de génération. Il peut combiner recherche sémantique et lexicale, filtres métiers et reclassement des résultats. Pour répondre à « quelle politique encadre ce seuil ? », la priorité est de retrouver le bon passage, dans sa version applicable et avec une citation vérifiable.

GraphRAG désigne une famille d'approches qui utilise des relations structurées pour enrichir la recherche. Dans la proposition de Microsoft Research, les documents servent à extraire des entités et leurs relations, à construire des communautés et à produire des résumés transversaux. Cette représentation aide à traiter des questions comme « quels problèmes récurrents relient nos projets, équipes et incidents ? ». Elle ajoute également des étapes de construction, de validation et de maintenance [1].

RAG

Rechercher puis générer avec des sources

Préparer le corpus, l'indexer, récupérer des passages, composer un contexte autorisé et demander au modèle de répondre. La chaîne peut aller bien au-delà d'une recherche vectorielle simple : recherche hybride, filtres d'attributs, reclassement, expansion de requêtes, citations et contrôle de versions.

Force : question précise dont le passage pertinent peut être retrouvé. **Risque :** omettre un fait éloigné, retenir une version périmée, restituer une information hors des droits.

GRAPHRAG

Expliciter des relations pour enrichir le contexte

Il n'existe pas une seule définition opérationnelle : construire un graphe d'entités et de relations, exploiter des graphes existants, rechercher des chemins relationnels ou résumer des communautés.

Force : relations, dépendances, synthèse transverse. **Risque :** des relations erronées ou obsolètes, présentées avec un vernis de précision et répercutées dans la réponse.

L'outil open source de Microsoft propose d'ailleurs plusieurs stratégies : Basic Search (référence RAG), Local Search (autour d'entités), Global Search (synthèse de communautés) et DRIFT Search (exploration locale éclairée par une vue globale). GraphRAG n'est donc pas une seule configuration universelle [4].

BASIC
Référence RAG vectorielle

LOCAL
Autour d'entités, avec passages bruts

GLOBAL
Synthèse des rapports de communautés

DRIFT
Exploration locale éclairée par une vue globale

02 / QUELLE APPROCHE POUR QUELLE QUESTION

Question professionnelle	Première approche à tester	Point de vigilance
Retrouver une clause, un contrôle ou une procédure	RAG hybride avec citations	Exhaustivité du passage, version, droits
Identifier les relations entre acteurs, obligations ou événements	Recherche guidée par graphe	Exactitude des relations, provenance
Synthétiser les tendances de centaines de dossiers	GraphRAG global ou synthèse hiérarchique concurrente	Couverture, diversité, coût de la synthèse
Combiner une vue d'ensemble et une enquête ciblée	Routage hybride ou mode de recherche mixte	Complexité, latence, erreurs de routage
Répondre à partir de règles actualisées régulièrement	Comparaison sous contrainte de fraîcheur	Réindexation, versions et retrait de faits obsolètes

Ces orientations sont des hypothèses d'architecture à valider sur les données de l'organisation, pas des résultats de tests réalisés par Spotwork.AI.

Cinq familles de questions et leurs preuves de réussite

Famille	Besoin et approche à évaluer	Preuve de réussite
Locale — quel seuil s'applique ?	Retrouver un passage exact, version en vigueur ; RAG hybride, filtres et reclassement	Citation exacte, couverture du passage, absence d'invention
Relationnelle — quels contrôles couvrent cette obligation ?	Traverser des liens documentés ; RAG enrichi d'un graphe ou recherche guidée	Exactitude de la relation, chemin traçable, source d'origine
Globale — quels thèmes émergent ?	Synthèse d'une collection ; GraphRAG global ou synthèse hiérarchique concurrente	Couverture des thèmes, diversité, exemples attribués
Comparative — différences entre filiales ?	Comparer entités et conditions ; routage mixte, graphe et passages bruts	Exhaustivité des dimensions, contradictions identifiées
Évolutive — qu'est-ce qui a changé ?	Documents datés et modifiés ; recherche temporelle et versionnée	Chronologie, versions et sources actives

Le choix d'une architecture doit être fondé sur la distribution réelle de ces questions. Si, par exemple, la grande majorité des usages porte sur des faits locaux, la valeur économique d'une chaîne globale complexe peut être faible, même si ses démonstrations sont spectaculaires.

03 Ce que démontrent — et ne démontrent pas — les recherches

La publication fondatrice de Microsoft (2024) montre des améliorations sur des questions globales de synthèse portant sur des jeux de données de l'ordre d'un million de tokens, notamment en couverture et diversité des réponses face à un RAG de référence. Le résultat concerne cette classe de tâches et son protocole : il ne démontre pas que tous les systèmes RAG sont inférieurs [1].

Une comparaison systématique indépendante publiée en prépublication en 2025 observe au contraire des forces distinctes selon les tâches et les métriques, et examine l'intérêt de combiner plusieurs approches. Son message utile pour les entreprises est de tester séparément recherche factuelle, raisonnement relationnel et synthèse [2].

Plus récemment, une étude publiée le 15 septembre 2026 compare des composants GraphRAG sur trois ensembles de données, dont un corpus de réglementation bancaire européenne. Dans une comparaison fondée sur 500 questions générées, sa configuration optimisée obtient un taux global de préférence de 60,6 % face à un RAG vectoriel de référence, évalué par un juge LLM sur plusieurs critères. Ce chiffre n'est ni un taux d'exactitude de 60,6 %, ni une garantie de performance sur d'autres corpus. Les auteurs notent des limites liées aux graphes statiques, à la langue et à l'évaluation automatisée [3].

Étude	Résultat principal	Limite pour un décideur
Edge et al., Microsoft (2024) [1]	Meilleure couverture et diversité que le RAG de référence sur des questions globales	Travaux de l'équipe conceptrice ; classe de tâches spécifique
Han et al. (2025) [2]	Avantages variables selon la tâche et la mesure ; intérêt d'approches combinées	Prépublication ; comparabilité sensible aux métriques, modèles et corpus
Liao et al. (sept. 2026) [3]	Préférence de 60,6 % pour la configuration GraphRAG optimisée, sur 500 questions générées	Jugement par LLM, questions générées, graphes statiques : ce n'est pas 60,6 % de réponses correctes
Microsoft Research (nov. 2024) [5]	−77 % de coût moyen en tokens pour la recherche globale dynamique, par rapport à la recherche globale statique de niveau 1, à qualité comparable, sur 50 questions	Comparaison entre deux variantes de recherche globale GraphRAG, dans cette expérience ; ni économie sur toute l'architecture, ni comparaison avec un RAG vectoriel

Conclusion : la structure des connaissances peut créer de la valeur, mais le bénéfice dépend du type de question, de la qualité du graphe et de la configuration retenue. Le terme GraphRAG agrège plusieurs techniques : les économies ou les améliorations constatées avec une variante ne se transfèrent pas mécaniquement aux autres.

04 La méthode Spotwork.AI : cinq décisions avant d'investir dans un graphe

Nous proposons une grille d'arbitrage, à éprouver sur chaque organisation. Elle ne constitue ni un benchmark propriétaire ni une norme scientifique.

Décision	Question à trancher	Preuve attendue
1. Qualifier	Quelles familles de questions les utilisateurs posent-ils réellement ?	Échantillon de questions locales, relationnelles et globales
2. Fiabiliser	Les documents, entités, relations et versions sont-ils exploitables ?	Sources identifiables, droits, fraîcheur, erreurs du graphe mesurées
3. Comparer	Une approche graphe améliore-t-elle le résultat par rapport à un RAG solide ?	Même corpus, mêmes questions, modèles et critères comparables
4. Arbitrer	Le gain justifie-t-il construction, mises à jour et temps de réponse ?	Coût complet par livrable accepté (ici, la réponse acceptée), latence, qualité et effort de revue
5. Gouverner	Comment détecter les dérives et corriger les connaissances ?	Journal des versions, tests de non-régression et responsables désignés

LE POINT DÉCISIF : UNE COMPARAISON LOYALE

Le point décisif est la troisième étape : ne pas opposer GraphRAG à un RAG volontairement simpliste. Une comparaison utile inclut une base RAG optimisée et, si le besoin le justifie, une variante relationnelle ou globale. Une démonstration qui compare un GraphRAG finement configuré à un RAG vectoriel rudimentaire informe sur la valeur du système complet, sans isoler l'effet du graphe. Garder fixes autant que possible le modèle, l'accès aux données, les critères et les budgets d'exécution. Une moyenne élevée ne compense pas une fuite de données ou une erreur juridique bloquante.

04 / MÉTHODE — SUITE

Protocole de pilote reproductible

Test	Comparaison proposée	Résultat attendu
T1 — Recherche locale	Même question factuelle, trois architectures	Source correcte et bonne version, sans invention
T2 — Raisonnement relationnel	Questions à plusieurs liens documentés	Relations exactes et traçables à un document
T3 — Synthèse globale	Questions multi-documents	Couverture et diversité, avec preuves citées
T4 — Contradictions	Deux sources en désaccord	Désaccord signalé, non fusionné sans justification
T5 — Fraîcheur	Document remplacé ou règle supprimée	Ancienne information non réutilisée après mise à jour
T6 — Sécurité	Accès différencié selon l'utilisateur	Aucun document ou lien interdit dans les réponses
T7 — Coût	Indexer, rafraîchir et interroger le même corpus	Coût complet et latence comparables

Proposition méthodologique, non une norme. Les cas sont revus et annotés par les métiers ; la taille de l'échantillon, les répétitions et les seuils d'acceptation sont arrêtés avant l'exécution, selon la complexité et le risque.

05 Gouvernance des connaissances : les points souvent oubliés

Provenance du document jusqu'à la réponse

Chaque réponse significative devrait pouvoir relier sa conclusion à des passages d'origine. Pour les relations de graphe, il faut en outre identifier le document qui fonde le lien, sa date, la méthode d'extraction et son statut de validation. Une relation « A dépend de B » sans provenance ni temporalité devient rapidement un risque métier.

Droits d'accès de bout en bout

Les contrôles d'accès doivent s'appliquer avant la récupération et la synthèse, y compris sur les relations déduites, les résumés de communautés et les caches. Le graphe peut révéler des liens entre objets qu'un utilisateur n'est pas autorisé à voir ; filtrer uniquement le passage affiché à la fin ne suffit pas. Ce principe est un contrôle de conception recommandé, pas une fonctionnalité automatiquement garantie par GraphRAG.

Versions, rafraîchissement et données contradictoires

Une modification réglementaire peut invalider une relation ou une synthèse pré-calculée. La revue des graphes doit donc comporter une stratégie de mise à jour incrémentale ou de reconstruction maîtrisée, des tests de non-régression et la possibilité d'identifier les composants affectés. L'étude de Liao et al. signale l'absence de prise en charge des graphes temporels et des mises à jour incrémentales dans son implémentation [3].

05 / GOUVERNANCE — SUITE

Coût complet et latence

Une comparaison doit distinguer coût d'ingestion initial, coût des mises à jour, coût de requête, latence, effort d'expertise humaine et coût de contrôle. Les gains de tokens observés dans certaines variantes de GraphRAG ne garantissent pas un coût total plus faible lorsque l'indexation des entités et communautés est élevée. Une unité décisionnelle utile est le coût complet par livrable accepté — ici, la réponse acceptée — à condition de définir une politique stable d'acceptation et un périmètre de coûts homogène pour toutes les architectures.

06 Exemple : analyser un référentiel de conformité

SCÉNARIO ILLUSTRATIF non résultat expérimental Spotwork.AI

Une direction des risques veut répondre à deux questions : « quelle version de la règle s'applique à ce contrôle ? » et « quelles obligations communes relient plusieurs dispositifs de surveillance ? ».

QUESTION 1 · LOCALE

La première question exige surtout de retrouver une source précise, dans sa version applicable.

QUESTION 2 · RELATIONNELLE

La seconde demande de relier des obligations, des contrôles et des entités à travers plusieurs documents : elle constitue un candidat pertinent à une recherche guidée par graphe.

Le protocole doit vérifier dans les deux cas la citation des passages d'origine, le respect des droits, les relations extraites, les informations contradictoires et l'actualisation après modification réglementaire. Une réponse plus complète n'est pas automatiquement une réponse juridiquement correcte : la validation métier reste nécessaire.

Deux autres applications professionnelles

Études stratégiques et conseil

Un cabinet compare des dizaines de missions pour retrouver les hypothèses récurrentes, les risques sous-estimés et les secteurs liés. Une synthèse globale peut être pertinente, mais elle doit montrer quels documents soutiennent chaque tendance et éviter de donner le même poids à des dossiers incomplets.

Recherche et innovation

Une équipe R&D suit l'évolution de plusieurs hypothèses scientifiques à travers articles, essais et commentaires. Une représentation relationnelle peut faciliter la découverte de connexions ; sa valeur dépend de la qualité des références, de la temporalité et de la revue humaine des interprétations.

07 Ce qu'un décideur doit retenir

Ne commencez pas par choisir une technologie. Commencez par caractériser les questions et le niveau de confiance attendu. Un RAG bien gouverné peut suffire à de nombreux besoins ; GraphRAG mérite un pilote lorsque la valeur dépend réellement des relations et des analyses transversales. La preuve se construit en comparant les résultats, les erreurs, le coût total et la capacité à maintenir les connaissances à jour.

C'est l'équilibre à tenir : évaluer les méthodes disponibles, et développer des connaissances professionnelles fiables, contextualisées et maîtrisées. **Construire la bonne intelligence, pas la plus complexe.**

08 Sources et limites

[1] Edge et al., Microsoft Research (2024) — From Local to Global: A Graph RAG Approach to Query-Focused Summarization.
Travaux fondateurs de l'approche de synthèse globale ; résultats produits par ses concepteurs.
arxiv.org/abs/2404.16130

[2] Han et al. (2025) — RAG vs. GraphRAG: A Systematic Evaluation and Key Insights.
Prépublication ; résultats conditionnés par les jeux de données et méthodes.
arxiv.org/abs/2502.11371

[3] Liao et al. (2026) — GraphRAG Best Practices: A Modular Evaluation Framework and Regulatory Case Study (15 septembre 2026).
Étude de composants, domaine réglementaire ; jugement LLM, limites de généralisation.
doi.org/10.3233/SSW260013

[4] Microsoft GraphRAG — Query Engine overview.
Documentation technique des modes de recherche.
microsoft.github.io/graphrag/query/overview/

[5] Microsoft Research (15 novembre 2024) — GraphRAG: Improving global search via dynamic community selection.
Comparaison interne entre variantes GraphRAG.
microsoft.com/en-us/research/blog/graphrag-improving-global-search-via-dynamic-community-selection/

Cette publication est une analyse documentaire. Aucun test comparatif de RAG et GraphRAG n'a été exécuté par Spotwork.AI pour produire les conclusions chiffrées ci-dessus. Les résultats sont attribués à leurs auteurs, avec les limites de leurs protocoles.

Spotwork.AI

L'IA professionnelle de confiance

Mieux travailler aujourd'hui,
construire votre intelligence de demain.

SPOTLAB · RESEARCH & INSIGHTS · ANALYSE P14

RAG vs GraphRAG : quelle architecture choisir pour les connaissances d'entreprise ?

Une matrice adaptée à la nature des questions et du corpus

© 2026 SPOTWORK.AI