

Veille littéraire et médiation numérique : de « Mauvais genres en bibliothèques » à « Bibliosurf »

Bernard Strainchamps

Documentaliste et concepteur de Bibliosurf.com

Depuis le début des années 2000, des projets indépendants de veille et de valorisation littéraire ont expérimenté de nouvelles formes de médiation, hors des circuits commerciaux et institutionnels classiques. Cet article présente la continuité entre le site Mauvais genres en bibliothèques, initié dans les années 2000, et le site Bibliosurf qui lui succède aujourd'hui dans un environnement enrichi par les outils du Web sémantique et de l'intelligence artificielle. L'accent est mis sur les outils et techniques mobilisés au service de cette médiation, réutilisables en bibliothèques publiques.

Introduction

À l'heure où l'intelligence artificielle ouvre de nouvelles perspectives pour la veille, Bibliosurf se présente comme un objet hybride, entre outil de veille, catalogue de bibliothèque ou de librairie et espace de médiation littéraire. Fondé dans un contexte d'explosion des contenus critiques en ligne, notamment via les blogs, les podcasts et les vidéos, le projet Bibliosurf cherche à structurer, synthétiser et rendre intelligible un corpus foisonnant et hétérogène.

Bibliosurf ne se contente pas d'être un simple agrégateur, il offre une démarche éditorialisée, enrichie par des métadonnées normalisées et des essais en traitement automatique du langage. Il s'inscrit dans une démarche visant à soutenir la lecture grâce à des outils numériques qui favorisent tant l'exploration émotionnelle que la diversité des contenus : créer des liens entre des œuvres, des auteurs ou des sujets qui semblent déconnectés, si ce n'est par l'impact qu'ils ont sur le lecteur.

À l'origine du projet : intentions et choix

Bibliosurf est né de la volonté de cartographier la réception critique des romans contemporains à partir des chroniques littéraires issues du Web. L'objectif initial était double : offrir un outil de veille éditoriale pour les professionnels du livre et proposer aux lecteurs une forme de navigation fondée à la fois sur

l'émotion ressentie et sur des parcours croisés entre genres, thématiques, éditeurs, auteurs ou périodes. La navigation invite à entrer dans les œuvres non pas par des critères prédéfinis, mais par résonance subjective. La dimension transversale permet de sortir des classements habituels et de passer facilement d'un ensemble de livres à un autre. Le projet se démarque par son indépendance, aussi bien dans ses choix éditoriaux que dans sa manière de fonctionner, avec une méthode artisanale mais soignée pour choisir les sources et organiser les contenus.

Architecture documentaire et sémantique

Le socle du projet repose sur une structuration fine des données, fondée sur des normes bibliographiques interopérables et des outils de requêtage avancés : chaque livre est lié à des métadonnées enrichies via des identifiants normalisés (ISBN pour l'édition, ISNI pour les auteurs, VIAF pour les autorités, Wikidata pour les liens vers le Web de données)¹, permettant une interopérabilité forte avec d'autres référentiels.

L'utilisation de l'API (interface de programmation d'application) de YouTube a aussi demandé un apprentissage : gestion des clés d'accès, des quotas et des données renvoyées au format JSON (JavaScript Object Notation).

Les métadonnées sont récupérées via des protocoles comme le SRU qui encapsule le format MARC dans du XML, ou via des requêtes SPARQL sur Wikidata. Chaque page HTML contient un bloc JSON-LD qui structure les données de manière sémantique exploitable par les moteurs de recherche et applications tierces.

Expérimentations algorithmiques

Depuis les débuts de Bibliosurf, de nombreux indicateurs sont calculés afin de suivre l'évolution de la réception critique : nombre d'avis collectés, identification des coups de cœur critiques, ou encore mesure

¹ Voir glossaire technique en fin d'article.

de l'impact d'un ouvrage sur le Web littéraire. Ces données ont posé les bases d'une approche quantitative de la veille littéraire, aujourd'hui prolongée par l'intégration de modules d'intelligence artificielle.

L'intelligence artificielle (IA) générative permet la synthèse des contenus : génération de résumés littéraires à partir d'extraits critiques, et la production de FAQ orientées vers la recommandation. Outre les résumés et FAQ, l'IA permet également de générer des portraits littéraires de villes à partir des intrigues géolocalisées, de caractériser le style d'un auteur à partir des avis des lecteurs, ou encore de dresser un profil synthétique d'un éditeur à partir des œuvres qu'il publie et des thèmes qui lui sont associés. Bibliosurf utilise différents modèles de langage (LLM) en fonction des tâches visées et des coûts de traitement. SpaCy est utilisé pour repérer les tournures négatives dans les phrases, BARThez pour générer des résumés générés en français avec une capacité de reformulation réduite. OpenAI est utilisé pour les résumés des avis ou des FAQ à tonalité subjective. Chaque modèle a ses forces : SpaCy pour sa rapidité et sa précision en NLP traditionnel, BARThez pour sa spécialisation francophone, OpenAI pour sa souplesse conversationnelle. Trouver le prompt idéal s'avère souvent plus complexe que d'écrire le code Python lui-même, d'autant plus que coder devient aujourd'hui accessible à presque tout le monde avec l'aide de l'IA générative.

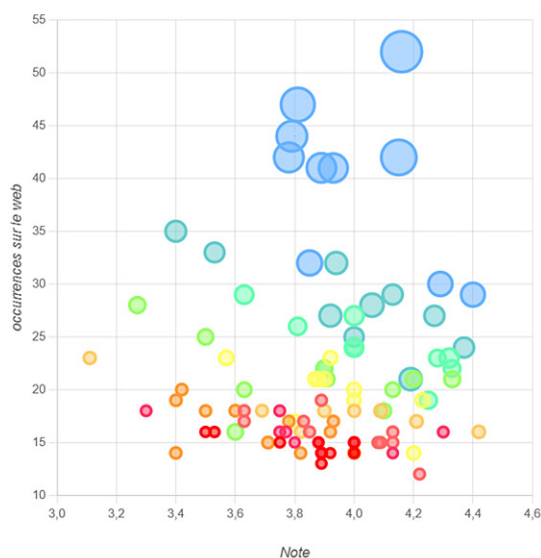
Médiation visuelle et exploration sensible

La dimension visuelle du projet repose sur plusieurs formes de datavisualisation, rendues possibles par l'usage combiné de bibliothèques comme Chart.js pour les graphiques interactifs, jqCloud pour les nuages de mots, Leaflet pour les cartographies. Bibliosurf constitue à ce titre un exemple assez unique dans le champ littéraire francophone d'un site mobilisant aussi largement les outils de visualisation de données au service de la médiation culturelle (voir figure). Des barres interactives représentent les émotions associées aux romans, des nuages de mots facilitent l'exploration thématique, des nuages de points visualisent l'impact de romans ou d'éditeurs sur une période donnée, et des cartographies littéraires permettent de découvrir des corpus géographiques (« Quand les livres racontent Marseille »).

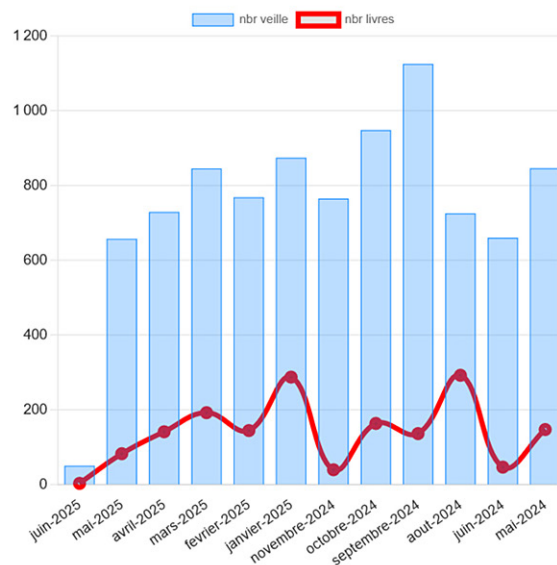
Ces visualisations, sobres dans leur esthétique – c'est-à-dire épurées, sans surcharge graphique, utilisant une palette de couleurs discrète et intégrées dans un design minimaliste – visent à mettre en valeur l'information sans détour. Elles sont également intuitives dans leur prise en main : faciles à comprendre sans mode d'emploi, elles permettent une interaction naturelle (cliquer sur une barre de données, un mot-clé, un point) et une exploration immédiate du contenu, sans nécessiter de compétences techniques particulières. Elles sont enfin optimisées pour tous les écrans, s'inscrivent dans une volonté de médiation augmentée : proposer une autre manière d'entrer dans les œuvres, fondée sur

Figure. Datavisualisation interactive à la une de Bibliosurf

Le palmarès des 12 derniers mois



La veille mois par mois



l'affect et la transversalité. Il existe aujourd'hui très peu de sites littéraires qui mobilisent aussi largement la datavisualisation pour structurer et enrichir l'accès aux contenus, ce qui confère à Bibliosurf une place singulière dans le paysage numérique francophone.

Usages et réception

Les statistiques fournis par Matomo révèlent une fréquentation en progression continue. Le cap des 1 000 visiteurs par jour est franchi. Les entrées sur le site s'effectuent principalement par les notices de livres et d'auteurs.

Après ces premières consultations, différents dispositifs invitent l'internaute à poursuivre sa navigation. Un travail de maillage interne favorise le rebond entre contenus. L'objectif est que le visiteur reste pendant plusieurs minutes. Chaque notice de livre intègre un algorithme de recommandation basé sur les avis des lecteurs. Des visualisations interactives (émotions, tags, cartographies) et des liens éditorialisés sont présents sur les pages auteurs, éditeurs ou collections. L'ensemble vise à prolonger l'exploration du site de manière intuitive et contextuelle. L'ajout d'accroches éditoriales (sous les barres d'émotion, les noms d'éditeurs, les tags) vise à encourager la navigation. Il convient également de souligner le travail éditorial mené sur les pages dites « pivots » du site : la page « Nouveautés en vue » anticipe les publications à paraître en fonction de leur potentiel critique, la « Revue du web » organise les retours critiques en temps réel, les « Portfolios d'auteurs » mettent en scène les trajectoires éditoriales, et les « Palmarès » agrègent les distinctions critiques sous différentes formes (classements mensuels, sélections thématiques, prix littéraires). Ces pages renforcent l'ancrage éditorial de Bibliosurf et constituent autant de points d'entrée alternatifs dans l'exploration du site.

Perspectives critiques

Bibliosurf s'inscrit dans les questions que pose aujourd'hui la médiation numérique : entre subjectivité et classement, entre lecture par l'humain et traitement par la machine, entre intervention éditoriale et automatisation. Les chroniques littéraires collectées en ligne, souvent marquées par l'émotion et le point de vue personnel, ne se laissent pas facilement enfermer dans des catégories fixes. Mais en parallèle, la nécessité de structurer les contenus pour en permettre l'exploitation documentaire impose un travail d'indexation, de normalisation et de catégorisation.

De la même façon, si la lecture humaine permet de saisir les nuances, les implicites et les contextes d'un texte critique, le traitement algorithmique, en revanche, permet d'agréger de grands volumes

d'informations pour en extraire des tendances, au risque parfois de simplifier ou de déformer le propos initial. L'enjeu est alors de trouver un équilibre entre puissance de calcul et fidélité interprétative.

Enfin, entre l'éditorialisation manuelle – assurée par le concepteur du site – et l'automatisation – permise par les outils d'intelligence artificielle – se pose la question du contrôle, de la cohérence éditoriale et de la responsabilité dans la mise en forme des contenus proposés.

Le projet explore les frontières entre documentation, médiation et recommandation littéraire. C'est une alternative aux logiques de prescription algorithmique dominantes. La lecture reste une expérience interprétative.

Conclusion

Bibliosurf constitue un laboratoire documentaire singulier. Ce site (sa méthodologie) pourrait être vu comme un exemple à la fois concret et prospectif, à enseigner dans le cadre des formations aux métiers des bibliothèques et de la documentation. Il mobilise un ensemble de compétences interdisciplinaires – en structuration de métadonnées, en visualisation de données, en analyse algorithmique des textes – qui sont au cœur des métiers de la documentation, de la veille, de l'édition numérique et de la médiation culturelle. Cette singularité tient à la combinaison rare d'approches hétérogènes dans un même projet : rigueur documentaire, subjectivité assumée, interopérabilité sémantique et recours ciblé à l'intelligence artificielle. Bibliosurf ne se contente pas d'appliquer des outils existants : il les détourne, les croise, les teste dans des configurations inédites, en privilégiant toujours la lisibilité, la nuance et l'autonomie éditoriale.

Bibliosurf est un terrain d'expérimentation, qui questionne les usages concrets de la veille littéraire et explore de nouvelles façons de la mettre en valeur. Sa structure ouverte, indépendante et sans but commercial lui donne une vraie liberté : dans les choix éditoriaux comme dans les essais techniques. En croisant rigueur documentaire, approche sensible et algorithmes, le projet propose une autre manière d'imaginer la médiation littéraire à l'heure de l'IA. ☺

Glossaire technique

- **Chart.js** : bibliothèque JavaScript pour produire des graphiques interactifs.
- **ISBN (International Standard Book Number)** : identifiant unique attribué à chaque édition d'un livre publié.
- **ISNI (International Standard Name Identifier)** : identifiant normalisé pour les noms d'auteurs, éditeurs ou institutions.

- **jqCloud** : plugin jQuery permettant de générer des nuages de mots dynamiques.
- **JSON-LD (JavaScript Object Notation for Linked Data)** : format léger pour structurer les données liées dans les pages Web.
- **Leaflet** : bibliothèque JavaScript légère pour la création de cartes interactives.
- **LLM (Large Language Model)** : modèle d'intelligence artificielle entraîné sur de vastes corpus de textes pour produire ou analyser du langage humain.
- **MARC (Machine-Readable Cataloging)** : format standard utilisé pour la description des ressources dans les bibliothèques.
- **Matomo** : outil open source de mesure d'audience web, utilisé pour analyser les comportements de navigation.
- **NLP (Natural Language Processing)** : ensemble des techniques permettant à un ordinateur d'analyser, de comprendre et de générer du langage naturel.
- **Prompt** : texte ou instruction fourni à un modèle de langage (LLM) pour générer une réponse ; sa formulation influe directement sur la qualité du résultat.
- **Python** : langage de programmation utilisé notamment pour le traitement de données, l'analyse de texte et le déploiement de modèles d'IA.
- **Schema.org** : vocabulaire standard pour l'annotation sémantique des données dans les sites Web.
- **SPARQL** : langage de requête permettant d'interroger des bases de données RDF comme Wikidata.
- **SRU (Search/Retrieve via URL)** : protocole normalisé pour interroger des catalogues bibliographiques à distance.
- **VIAF (Virtual International Authority File)** : agrégateur d'autorités bibliographiques internationales.
- **Wikidata** : base de connaissances collaborative structurée, interopérable avec d'autres projets comme Wikipédia.