La numérisation du patrimoine. Du projet Gutenberg à Google Arts & Culture

Marie Puren

LA NUMÉRISATION DU PATRIMOINE

Du projet Gutenberg à Google Arts & Culture

Marie Puren

16 octobre 2020 : CM1 supplémentaire

UVSQ | M1 ACPCI, ECMAH, EMAS, HCS, RCL

PLAN

INTRODUCTION

LES DÉBUTS

- Etats-Unis : nouvelles technologies vues comme une possibilité viable de donner plus largement accès aux collections patrimoniales.
- 1990 : la Bibliothèque du Congrès avec le projet pilote “American Memory” : http ://memory.loc.gov/ammem/browse/updatedList.html.
- France pionnière pour la numérisation du patrimoine avec Gallica

LES BIBLIOTHÈQUES NUMÉRIQUES

DÉFINITION

“Les bibliothèques numériques proposent de véritables collections numériques, selon une politique documentaire déterminée. Elles sont alimentées soit par des opérations de numérisation (documents patrimoniaux ou non), soit par des documents nativement numériques. Les contenus sont organisés pour en faciliter la consultation”.

Définition de l’Ecole nationale supérieur des sciences de l’information et des bibliothèques (Enssib)

Le projet Gutenberg, 1ère bibliothèque numérique

- Projet Gutenberg considéré comme la première bibliothèque numérique. Créée par Michael Hart en 1971 à l'Université de l'Illinois.Soixantaine de langues représentées, pour 54.000 livres électroniques en accès libre.
- Projet Runeberg lancé en décembre 1992 pour la littérature scandinave
- Projet Gutenberg-DE en 1994 pour la littérature allemande.
- Deuxième grand projet original de bibliothèque numérique plus tardif : The Online Books Page, initié en 1993 par John Mark Ockerbloom.
- Pour les bibliothèques numériques françaises, projet précurseur : bibliothèque électronique de la bibliothèque municipale de Lisieux (1996).

LES 1ÈRES BIBLIOTHÈQUES NUMÉRIQUES FRANÇAISES

- 1997 : Gallica = rendre accessibles en ligne des documents libres de droit.
- L'idée d'une bibliothèque numérique nationale remonte au 14 juillet 1989. François Mitterrand : création d'une "bibliothèque d'un genre nouveau".
- Première version lancée en 1997 et dernière version de la plateforme en 2015.
- En chiffres : plus de 3 millions de numéros de presse ou de revues, près de 1.400.000 images ou encore 660000 livres.
- Initiatives de la part de bibliothèques municipales : exemple de Rouen, dès 2001.

Accélération dans les années 2000

- Europeana à partir de 2005
- Bibliothèque numérique mondiale, lancée en décembre 2006 par l'UNESCO et la Bibliothèque du congrès
  - Projet ambitieux, qui a pour but de réunir sur Internet des documents représentatifs de la culture mondiale de tous les pays
  - Lancée en 2009 ; à ce jour, 158 partenaires dans 60 pays

DEUX GRANDS PROJETS CONCURRENTS

Google Livres : la 1ère initiative d'envergure

- Lancement de Google Book Search [1] lors de la foire du livre de Francfort en octobre 2004.
- A partir de 2006, partenariats avec cinq grandes bibliothèques de recherche anglo-saxonnes : la New York Public Library, la bibliothèque de Harvard, la bibliothèque de l'Université du Michigan, la bibliothèque de Stanford et la Bodleian Library d'Oxford.
- 800.000 ouvrages numérisés pour la seule bibliothèque d'Harvard.

1. Pour plus d'informations sur Google Livres, on peut consulter : Bruno Racine, Google et le nouveau monde, Perrin, 2011.

FONCTIONNEMENT

- Prend en charge les frais inhérents à la numérisation et à la conversion en mode texte des ouvrages et fournit en retour les fichiers numériques
- En contrepartie, droit d'exploiter commercialement ces ouvrages pour une durée de 25 ans.
- Obligation de limiter, pendant la même durée, l'exploitation de ces ouvrages par un potentiel concurrent de Google.

UN PROJET INTERNATIONAL

- Signature de conventions avec de nombreuses bibliothèques européennes et japonaises, comme la Complutense de Madrid, la Keiõ de Tokyo, la bibliothèque cantonale de Lausanne et la Bibliothèque municipale de Lyon.
- 2015 : 25 millions d'ouvrages, dans plus de 400 langues.

DES INFRACTIONS AU DROIT D'AUTEUR

- Septembre 2005 : poursuite de l'Authors Guild (le syndicat des auteurs américains) pour violation du droit d'auteur
- Suivi d'un procès intenté par l'association des éditeurs américains
- 2015 : jugement de la cour d'appel de New York en faveur de Google
- France : procès avec certains éditeurs comme La Martinière (2006-2009) ; accord avec Hachette Livre (2011)

UTILISER GOOGLE LIVRES

- Un moteur de recherche simple
- Interface de recherche avancée

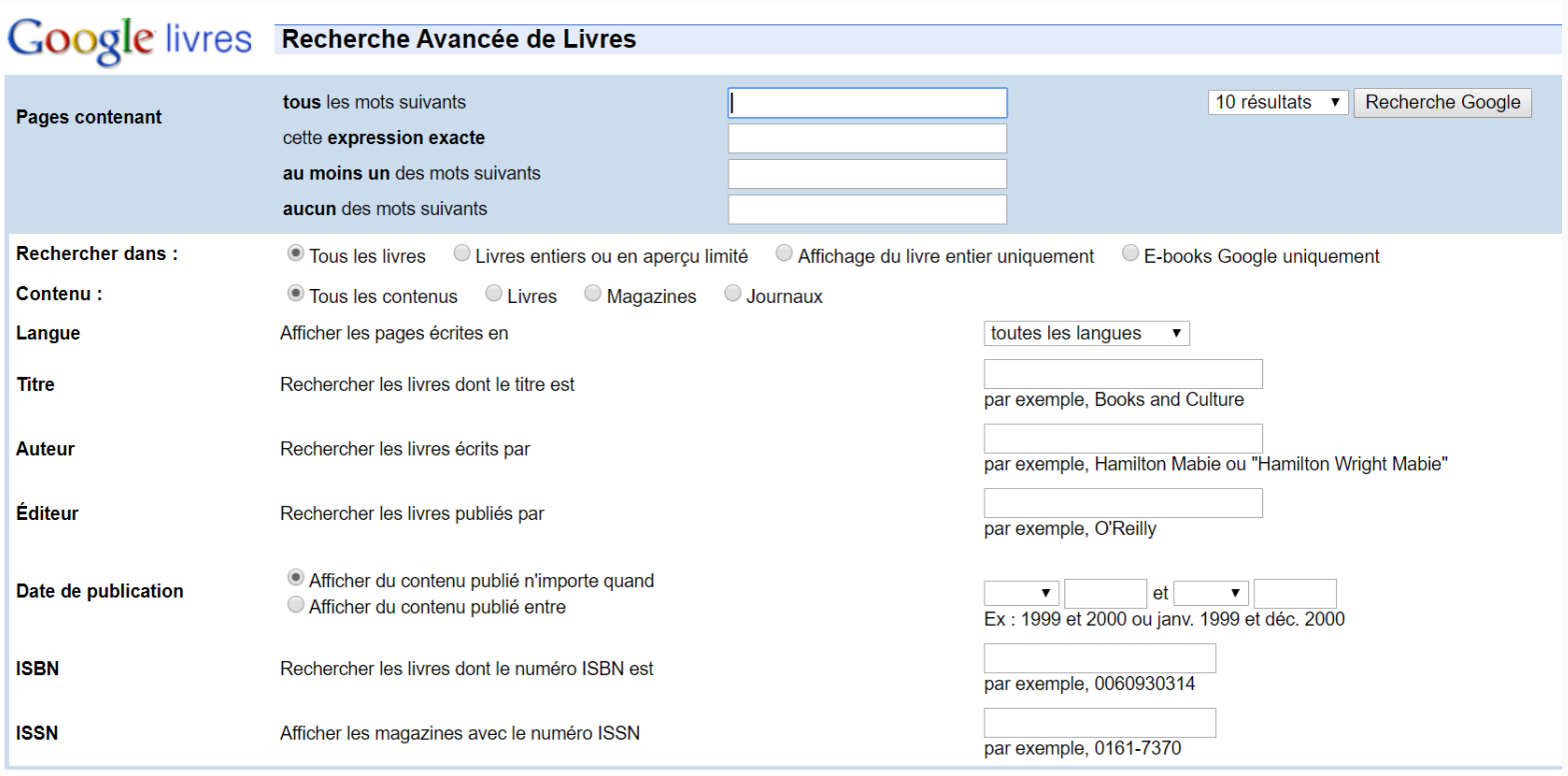


FIGURE : Recherche avancée - Google Livres

CRITIQUES

- Jean-Claude Guédon : perspective effrayante que de donner autant de contrôle à une seule entreprise sur la mémoire collective mondiale, son analyse et même sa signification [2].
- Robert Darnton : les livres sont des biens communs, et seules des organisations publiques, sans but lucratif, devraient avoir le pouvoir de les contrôler [3].
- Création de l'OCA ou Open Content Alliance en Octobre 2005, terminé en mars 2011. 300.000 livres numérisés disponibles dans l'Internet Archive.

2. Jean-Claude Guédon, "Who will digitize the world's books ?", dans The New York Review of books, 4 août 2008
3. Robert Darnton, "The Library in the New Age", dans Ibid., 12 juin 2008

GOOGLE NGRAM VIEWER

- Décembre 2011 : Ngram Viewer par Google.
- Visualiser la fréquence d'apparition d'une suite de mots, sous la forme de courbes, dans les livres numérisés par Google Livres.
- **"ngram" = suite de "n" mots**
- Corpus lexical basé sur les livres numérisés par Google Livres. Sous-lexiquez par langues : Français, Anglais américain ou britannique, Espagnol, etc.
- Lexiques = tables composés de n-grammes, c'est-à-dire des séquences de mots qui apparaissent dans les ouvrages numérisés.
- 5 catégories de table : monogrammes (un seul mot), bigrammes (deux mots qui se suivent) et ainsi de suite jusqu'à une suite de 5 mots.

GOOGLE NGRAM VIEWER

Impossible de connaître la fréquence d'apparition de ce vers de Phèdre "ma folle ardeur malgré moi se déclare" (acte II, scène 5), mais de "ma folle ardeur malgré moi" et "ardeur malgré moi se déclare".

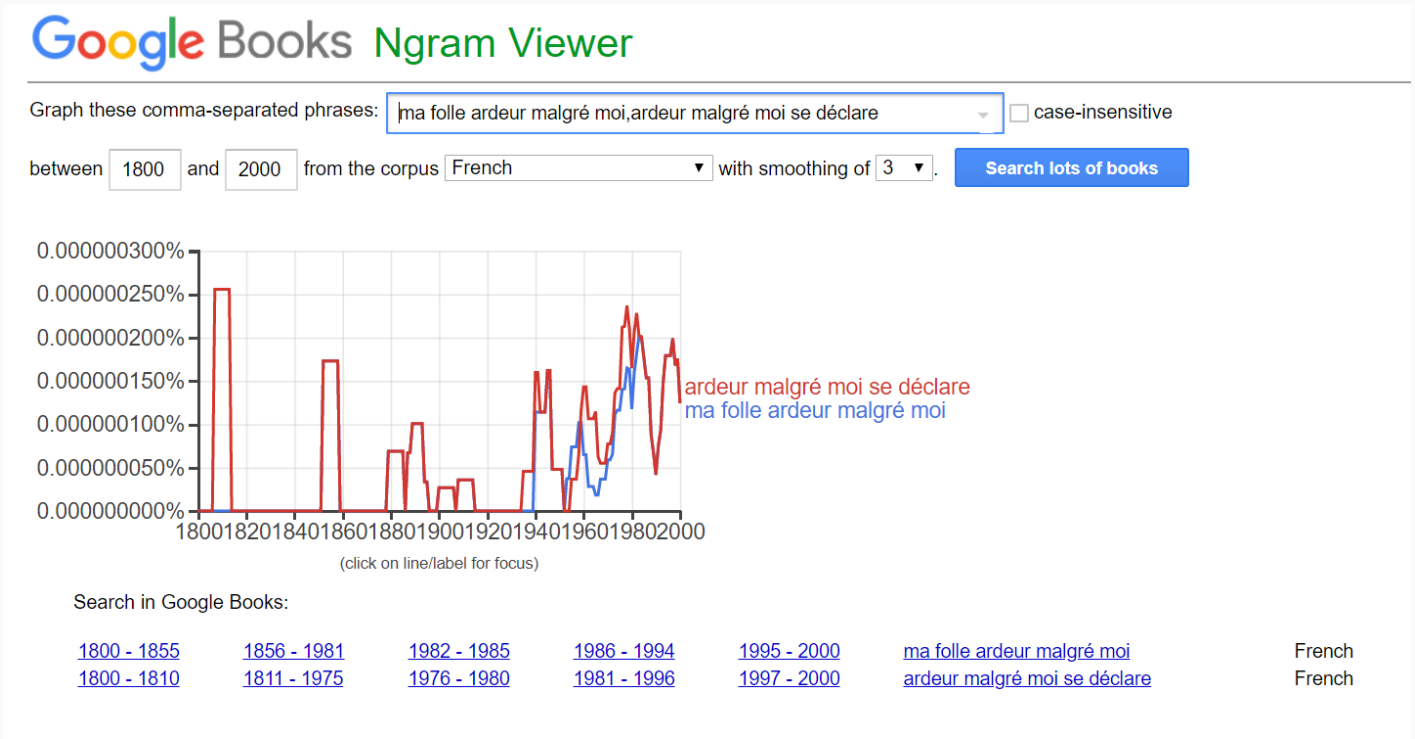


FIGURE : Ngram Viewer

GOOGLE NGRAM VIEWER

- Impossible d'interpréter correctement ces graphiques sans avoir accès aux documents numérisés sur lesquels se basent ses résultats.
- Tâche colossale, au vue de l'ampleur du corpus numérisé
- Difficilement réalisable car beaucoup de documents sont encore sous droits et pas consultables.
- Ngram viewer : outil heuristique, destiné à poser des questions et à déceler des tendances.

EUROPEANA : UNE RÉPONSE À GOOGLE LIVRES

- Début Europeana en septembre 20025
- Conçu comme le concurrent direct de Google Books

EUROPEANA : UNE RÉPONSE À GOOGLE LIVRES

“Voici que s’affirme le risque d’une domination écrasante de l’Amérique dans la définition de l’idée que les prochaines générations se feront du monde. [...] Toute entreprise de ce genre implique [...] des choix drastiques, parmi l’immensité du possible. Les bibliothèques qui vont se lancer dans cette entreprise sont certes généreusement ouvertes à la civilisation et aux œuvres des autres pays. Il n’empêche : les critères de choix seront puissamment marqués [...] par le regard qui est celui des Anglo-Saxons, avec ses couleurs spécifiques par rapport à la diversité des civilisations”.

Jean-Noël Jeanneney, “Quand Google défie l’Europe”, dans Le Monde le 23 janvier 2005

LA MAÎTRISE DE L'INFORMATION NUMÉRIQUE

- Information numérique = enjeu de politique internationale.
- Philosophe Jean-François Lyotard en 1979 : "Comme les États-nations se sont battus pour maîtriser des territoires, puis pour maîtriser la disposition et l'exploitation des matières premières, il est pensable qu'ils se battent à l'avenir pour maîtriser des informations. Ainsi se trouve ouvert un nouveau champ pour les stratégies industrielles et commerciales et pour les stratégies militaires et politiques [4]".

4. Jean-François Lyotard, La Condition postmoderne, Paris, Les Éditions de Minuit, 1979, p. 15.

AMBITIONS ET ÉVOLUTION D'EUROPEANA

- Europeana => totalité du patrimoine culturel européen numérisé.
- Site lancé le 20 novembre 2008.
- 2016 : Europeana Collections.
- 48 millions d'oeuvres d'art, d'objets, de livres, de vidéos ou de sons issues des collections numérisées de plus de 3300 institutions européennes.

D'AUTRES PROJETS DE BIBLIOTHÈQUES NUMÉRIQUES

- 2005 : lancement de Persée (Portail de Revues en Sciences Humaines et Sociales) par le Ministère de l'Enseignement Supérieur et de la Recherche. Destiné à compléter l'offre de Gallica en se concentrant sur les publications scientifiques.
- Bibliothèque nationale de Norvège : très vaste programme de numérisation en masse => numériser tous les documents qu'elle conserve : livres, images, sons...etc. Fin prévue de ce chantier dans 20 ou 30 ans.

D'AUTRES EXEMPLES

- Images Online : banque iconographique de la British Library
- Les classiques des sciences sociales : texte intégral de grands auteurs classiques en sociologie, anthropologie, économie, sciences politiques, philosophie sociale et politique
- Manuscriptorium : diffusion de ressources historiques sur le livre comme des manuscrits, des incunables, des cartes, des chartes...etc.
- Atlas : base de données des oeuvres exposées au Louvre
- Collection du Centre Pompidou
- Collections numérisées de la bibliothèque de l'INHA

D'AUTRES PROJETS DE NUMÉRISATION : EXEMPLES

LA NUMÉRISATION DES ARCHIVES

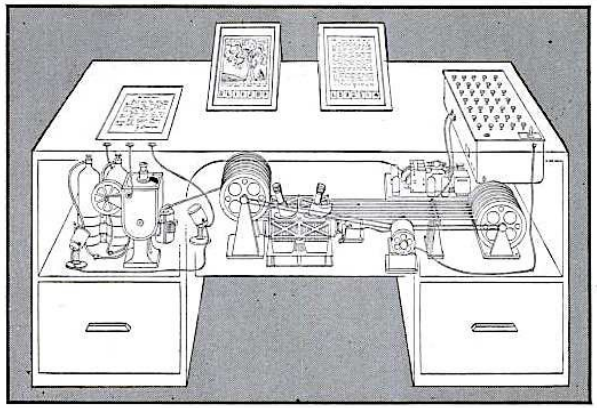
- Début de la numérisation des archives en France dans le milieu des années quatre-vingt-dix.
- Archives départementales, mais aussi archives municipales : premiers services à voir l'intérêt de ces grandes campagnes de numérisation.
- Liste des fonds d'archives numérisés sur Patrimoine numérique qui est le catalogue collectif national des collections numérisées en France.

MYLIFEBITS

- MyLifeBits : projet de recherche initié par Microsoft en novembre 2001. Orojet d'archivage total d'une vie, entrepris par Gordon Bell.
- Stocker, de manière automatique, les documents, les images, les sons - comme les conversations -, et d'y avoir accès facilement et rapidement.

MYLIFEBITS

- Réaliser la vision de Vannevar Bush qui a imaginé le concept du memex, un ordinateur fictif décrit en 1945 dans un article intitulé “As We May Think”.



MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

AS WE MAY THINK CONTINUED

FIGURE : Le Memex

- Logiciel MyLifeBits développé pour le projet par Jim Gemmel et Roger Lueder. Décrit dans ”Total Recall : How the E-Memory Revolution Will Change Everything” par Gordon Bell et Jim Gemmel en 2009.

GOOGLE ARTS & CULTURE : L'ART NUMÉRISÉ

UN AUTRE PROJET MENÉ PAR GOOGLE

- Février 2011 : Google Art Project devenu Google Arts & Culture
- Plateforme en ligne qui donne accès à des images en haute-résolution d'oeuvres d'art conservés par des musées et institutions partenaires.

CRITIQUES

- Réutilisation des images : bien que les oeuvres soient elles-mêmes libres de droit, pas possible de télécharger ces images.
- Pas le droit de faire autre chose que de consulter ces images en ligne.
- Catégorie “Google Art Project” de Wikimedia Commons donne accès, en téléchargement, à des images provenant du Google Art Culture.
- Cf. article de 2011 dans Libération : images reconstituées par un programme informatique

LA TOTALITÉ DU PATRIMOINE ACCESSIBLE ?

LE PATRIMOINE NUMÉRIQUE

- Association des termes “patrimoine” et “numérique” pour la première fois dans un texte de l’Unesco en 2003, la Charte sur la conservation du patrimoine numérique
- Idée que la numérisation allait permettre de mieux conserver les collections patrimoniales, et de faciliter la transmission des savoirs et des connaissances.
- Politiques nationales et internationales surtout tournées vers la numérisation de masse, laissant de côté une sélection plus qualitative des documents numérisés.

LE MYTHE DE L'ACCÈS UNIVERSEL AU PATRIMOINE

- Illusion quant aux potentialités du Web pour l'accès au patrimoine.
- Web souvent perçu comme une source de savoir exhaustive, alors que la majeure partie du patrimoine mondial n'est pas accessible par le biais des nouvelles technologies.
- Patrimoine pas numérisé, ou difficilement accessible à cause du manque de données pour le décrire ; ou restrictions dues à la protection du droit d'auteur et de la vie privée.
- Pas oublier que le choix de numériser ou non un document patrimoniale dépend des financements reçus
- Seulement une petite partie du patrimoine mondial numérisé.