Les humanités pour le numérique ?

Thierry Daunois[1]
[1] Université de Lorraine, Direction des partenariats, Vandoeuvre-lès-Nancy, France
thierry.daunois@univ-lorraine.fr

Abstract :
Cet article propose, dans la lignée et en contrepoint des travaux menés dans le champ des humanités numériques - qui visent à mettre les méthodes du numérique au service des sciences humaines et sociales -, de réfléchir à la façon dont ces dernières pourraient constituer une approche nouvelle et pertinente aux problématiques de la fouille de données et de la préservation des données numériques pour les chercheurs de demain. L'accent sera mis plus particulièrement sur l'intérêt de mobiliser les outils et les méthodologies de l'archéologie.

Mots-clés : web, ingénierie de la connaissance, humanités numériques, *digital humanities*, archéologie, *data mining*, *web archaeology*, archéologie du web

1. Introduction

On voit fleurir, depuis quelques années, réflexions et recherches sur l'histoire de l'informatique, et, plus précisément, l'histoire d'Internet et du web (306 000 réponses à la requête “histoire du web”, 89 000 pour “histoire de l'Internet”, et 56 000 pour “histoire d'Internet” sur Google, à la date du 9 septembre 2013). L'Internet Society propose, sur son site, un document conséquent sur le sujet, parmi bien d'autres [1]. Dans la plupart de ces démarches, sont proposées des dates, des faits, des événements (création d'ARPANET, protocoles, acteurs...). On trouve également des documents se rapportant à l'usage que peuvent avoir les sciences humaines et sociales du web, avec le développement rapide des *digital humanities* [2]. Plusieurs livres et articles scientifiques font ainsi référence à une “archéologie électronique” [3,4].

On voit également se développer des réflexions autour d'une philosophie du web, notamment autour des travaux, en France, d'Alexandre Monnin [5,6], parallèlement au milieu “hacker”, où une “archéologie du code” est désormais évoquée, autour du legacy code. Comment les développeurs, confrontés à d'anciens programmes doivent et peuvent-ils les faire évoluer au rythme des nouvelles plateformes ? Cette réflexion préfigure probablement la difficulté à laquelle seront confrontés tous ceux qui, dans quelques années, travailleront sur du “vieux” code [7].

2. Archéologie et recherche d'informations sur le web : analogies

Toute personne utilisant régulièrement Internet comme source d'informations peut observer des rapprochements, au moins sémantiques, avec l'archéologie. On parle de “fouille” (site de fouille et fouille de données), on se segmente par “langages” (dialectes du grec ancien, assyrien, ou langages informatiques obsolètes, comme APL ou Algol), on se confronte à des “traces”, ou “vestiges” dont seul le contexte donne la clé (le concept de “strates” s'applique aussi bien à un terrain matériel qu'immatériel, et

la mise en scène télévisuelle de la police scientifique a mis en évidence l'idée de “traces informatiques”).

Ces concordances de vocabulaire pourraient, certes, n'être que mimétisme : pour décrire des opérations nouvelles, les chercheurs emploient souvent des termes se référant à des champs disciplinaires préexistants, au risque de l'abus de langage. Mais le parallèle peut être poursuivi : avec la fermeture de l'une (sinon de la première, les historiens du web sauraient le dire !) des premières plateformes d'hébergement gratuites, GéoCities, en 2009, on fait face à l'engloutissement d'un “monde”. Créé en 1994, avant les moteurs de recherche comme Google, GéoCities formait un continent, sur lequel existaient plus d'une trentaine de “villes”, par thématique : Hollywood regroupait les amoureux du cinéma, Paris ceux de la poésie, Silicon Valley les accros de l'informatique... Lors de sa fermeture par Yahoo, un groupe d'activistes, The Archive Team, en a réalisé une ultime sauvegarde, incomplète. Qui, mieux que des archéologues, pourrait désormais explorer ce continent, cette première “web-Atlantide”, fût-elle numérique ?

Dans l'une et l'autre activité, il s'agit de travailler sur un terrain où, par définition, des éléments (voire des chaînons) sont manquants, avec une datation imprécise. Les archéologues disent qu'ils “fouillent les poubelles” [8] ; les “poubelles du web” (liens cassés, pages isolées, modifications...) peuvent être tout aussi explicites ! On peut alors s'interroger : les méthodes de l'archéologie, élaborées et structurées au fil des années [9] ne constitueraient-elles pas une bonne façon d'aborder, de structurer, voire d'ouvrir de nouvelles perspectives à l'activité de recherche d'informations ? Ne peut-on réfléchir sur l'apport des humanités au numérique, en inversant le paradigme ?

L'objet de cette réflexion – à un stade très préliminaire – est d'approfondir ce parallèle. N'est-il que de façade ? Ou dissimule-t-il des liens plus forts, l'archéologie pouvant alors proposer à la fois une grille d'analyse de l'activité de recherche d'informations sur Internet, et une source de bonnes pratiques pour la conception de sites Internet. Une telle démarche a déjà été proposée [10], c'est à sa systématisation que nous appelons. Que pourrait donc être une “archéologie du web”, et quelles pourraient être son(ses) utilité(s) et sa place ?

3. L'archéologie du web, nouvel angle de la fouille de données ?

La fouille de données (data mining) mobilise des “méthodes et algorithmes issus [notamment] des statistiques”, constituant un espace essentiellement occupé par des informaticiens [11]... dans lequel les utilisateurs ne se retrouvent pas toujours... [12].

La fouille de données nécessite de pouvoir réunir a priori des masses de données reliées au sujet choisi. Cette démarche, naturellement, ne peut pas être celle de l'archéologue, qui doit forcément travailler sur les données disponibles sur son site de fouille, qu'elles soient suffisantes ou non. La confrontation de ces deux modes de fonctionnement pourrait être extrêmement riche !

En revanche, on peut établir un autre parallèle entre ce qui est de l'ordre des “déchets” en archéologie (terrains décapés...) et le “bruit” en fouille de données, données signifiantes mais inexploitables dans le contexte concerné.

4. **L'archéologie du web, bonne pratique des concepteurs web ?**

L'idée de réfléchir en terme d'archéologie du web permet également de se projeter dans une direction radicalement différente. La question devient alors : quels seront les besoins de toutes celles et ceux qui voudront, à un moment ou un autre, analyser une part de la production web. Il est très peu douteux que, d'ici quelques années, les descendants des historiens qui, aujourd'hui, cherchent à décrypter les réseaux qui existaient au sein des sociétés savantes au XIXe siècle, auraient grande utilité de pouvoir travailler sur le contenu du web d'aujourd'hui.

Dès à présent, les forums Internet regorgent d'interrogations sur “comment puis-je identifier la date de création d'une page web ? La date de sa dernière modification ? Et les réponses font apparaître que ce n'est pas trivial. Or ces interrogations seront probablement celles de certains chercheurs dans cinquante ans. Une analyse détaillée des mécanismes déjà existants sera naturellement nécessaire.

En adoptant dès aujourd'hui la posture d'un archéologue du web, il est possible de déconstruire la fatalité qui veut que des éléments utiles doivent disparaître, simplement du fait que nos capacités de stockage des données numériques sont largement insuffisantes pour l'ensemble des données produites.

5. **Un exemple : si l'on datait au carbone 14 les objets numériques ?**

Il ne s'agit pas forcément d'un exemple réel et réaliste, mais d'une illustration de la démarche proposée ici. Rappelons que cette proposition en est à un stade préliminaire et qu'elle n'a pas été validée formellement.

La datation au carbone 14 est la plus médiatisée des méthodologies et techniques de l'archéologie – peut-être parce qu'elle a valu à son inventeur, Willard Franck Libby, le prix Nobel de chimie en 1960. Serait-elle “transposable” aux objets numériques ?

Définition de la datation au carbone 14 : “Cette méthode repose [...] sur le cycle de vie d'un des isotopes du carbone : le carbone 14 dit radiocarbone. [...]. Il est absorbé par les matières vivantes et stocké tout au long de la vie de l'individu. Lorsque la mort survient, le carbone 14 va progressivement disparaitre selon un schéma bien précis et désormais maitrisé” [13]. Grâce à ce schéma, on peut aujourd'hui déterminer assez précisément à quelle période l'être en question est mort.

Est-il envisageable de transférer cette démarche à des objets numériques, et ce, sous les deux angles de vue proposés précédemment ? Autrement dit, dans une optique de “fouille de données”, une partie du “code” qui définit l'objet numérique évolue-t-il au fil du temps, et selon un rythme qui serait suffisamment “stable” pour tenir lieu d'échelle de datation (on pense ici plus spécialement aux modes de stockage, les 0 et les 1 n'étant pas a priori sujets à évolution) ? Et, dans une optique de “bonnes pratiques de la conception de sites Internet”, est-il possible - et légitime - d'intégrer dans un langage informatique une part variable destinée à faciliter cette datation, pour l'avenir, y compris à une époque à laquelle ce langage ne serait plus “lisible” ?

6. **La démarche proposée**

La démarche repose sur la constitution d'un réseau international pluridisciplinaire

(archéologues, informaticiens, “digital humanists”...). Son premier rôle consistera à valider une liste initiale d'outils et de méthodes issues de l'archéologie, qui seront étudiés sous l'angle de leur applicabilité à la problématique de la fouille de données, ainsi qu'il est montré rapidement dans l'exemple ci-dessus. Dès que des outils potentiellement intéressants à transférer de l'archéologie à la recherche d'information sur Internet auront été identifiés, un terrain d'expérimentation sera mis en place. L'objectif sera, sur une question relevant de la fouille de données, de comparer les résultats qualitatifs et quantitatifs obtenus par les deux démarches et de mesurer s'il y a “recoupement” entre les données recueillies, pour déterminer si l'une des méthodes exclut l'autre, ou si elles sont complémentaires.

Ce travail doit constituer le coeur d'une thèse de doctorat. Les résultats attendus sont de trois ordres : constitution du réseau international pluridisciplinaire décrit plus haut ; les résultats d'un premier terrain d'expérimentation, pour valider ou invalider la démarche proposée ; l'identification de “bonnes pratiques” pour les concepteurs web travaillant avec les milieux académiques.

7. Conclusion

De nombreux éléments, et quelques réflexions publiées précédemment, forment un faisceau concordant qui semble souligner l'intérêt qu'il pourrait y avoir à observer le développement - et la structuration - du web sous l'angle des méthodologies de l'archéologie. Une telle démarche peut ouvrir d'une part sur une nouvelle façon d'envisager la fouille de données, mais probablement, plus largement, sur l'archéologie de la connaissance au sens foucaldien du terme [14].

Références

1. Un bref historique de l'Internet (Internet Society). Consulté en ligne (30 avril 2014), `http://www.internetsociety.org/fr/internet/qu%E2%80%99est-ce-que-l%E2%80%99internet/histoire-de-l%E2%80%99internet/un-bref-historique-de-linternet`.
2. On consultera notamment le I. 3. du Manifeste des Digital Humanities. Consulté en ligne (30 avril 2014), `http://tcp.hypotheses.org/318`.
3. Nassera Zaïd. "Archéologie électronique : La publication archéologique sur Internet". Dans : Archeologia, 1999, n°352, pp. 20-29. Consulté en ligne (30 avril 2014), `http://cat.inist.fr/?aModele=afficheN&cpsidt=2024873`.
4. Julien Mahoudeau. "Perspectives de recherche et d'action pour la valorisation scientifique : sur quelques expériences hypermédias en archéologie". Dans : Nouvelles perspectives en sciences sociales : revue internationale de systémique complexe et d'études relationnelles, Vol. 3, numéro 1, septembre 2007, p. 9-46. doi : 10.7202/602464ar. Consulté en ligne (30 avril 2014), `http://www.erudit.org/revue/npss/2007/v3/n1/602464ar.pdf`.
5. Alexandre Monnin, thèse Vers une philosophie du Web : le Web comme devenir-artefact de la philosophie (entre URIs, tags, ontologie(s) et ressources). Consulté en ligne (30 avril 2014), `http://tel.archives-ouvertes.fr/tel-00879147`.
6. Harry Halpin, Alexandre Monnin. Philosophical Engineering: Toward a

Philosophy of the Web, 2013, Wiley-Blackwell, 216 p.
7. Une définition possible du legacy code : "Selon The Economics Of Software Maintenance In The 21 Century, nous passons 80% de notre temps à maintenir du code existant et pénible à modifier. Dans cette situation nous devenons des archéologues du code [...]. On parle alors de code Legacy", sur le site des conférences Agile, http://2013.conference-agile.fr/sessions/indiana-jones-et-le-temple-du-legacy-code.html#.U2DLhsdnMzV. Michael Feathers (Working Effectively with Legacy Code) a redéfini le legacy code comme du “code sans test, ce qui le rendra difficile d’utilisation à l’avenir, notamment vis-à-vis des régressions”, sur le blog de la société So@t, http://blog.soat.fr/2013/05/agile-france-2013-code-legacy-faire-evoluer-ou-reecrire/. Sources consultées en ligne (30 avril 2014).
8. René Treuil. "L’archéologue amateur de déchets". Dans M. Tabeaud, Gr. Hamez (éds), Les métamorphoses du déchet, Paris, Publications de la Sorbonne, p. 69-73. Consulté en ligne (30 avril 2014) (Google books). http://abrege.eu/TreuilDechets.
9. Demoule Jean-Paul, Giligny François, Lehöerff Anne, Schnapp Alain. Guide des méthodes de l'archéologie, Ed. La Découverte 2009 (3e édition), 336 p.
10. Brysbaert Ann, Bocchi Laura, Tuosto Emilio. “Relating archaeological chaîne opératoire and process mining in computer science”. Dans : Archeologia e Calcolatori 23: 77-98, 2012. Consulté en ligne (3 juillet 2014), http://eprints.bice.rm.cnr.it/6098/1/10_Brysbaert_et_al.pdf.
11. Voir à ce sujet la présentation du cours sur la fouille de données de l'INSA de Rouen. Consulté en ligne (30 avril 2014), https://moodle.insa-rouen.fr/course/view.php?id=92.
12. Paragraphe "Les barrières majeures aux Data Mining efficaces" (page Data mining et systèmes d'aide à la décision) - ENST - 2000. Consulté en ligne (30 avril 2014), http://perso.telecom-paristech.fr/~saglio/bdas/00/exos/park/index.htm#_Toc477965838.
13. Définition sur le site archeologia.be. Consulté en ligne (30 avril 2014), http://www.archeologia.be/ressources_lexique_carbone-14.html.
14. Voir le commentaire d'Alexandre Serres sur L'archéologie du savoir. Consulté en ligne (30 avril 2014), http://hal.archives-ouvertes.fr/docs/00/10/84/51/PDF/Texte_ArcheologieSavoir.pdf.