



HAL
open science

Quels problèmes pour la recherche d'information médicale au Mali ?

Seydou S. Doumbia, Lorraine Goeuriot, Marie-Christine Fauvet

► To cite this version:

Seydou S. Doumbia, Lorraine Goeuriot, Marie-Christine Fauvet. Quels problèmes pour la recherche d'information médicale au Mali ?. RJCRI - CORIA, 2017, Marseille, France. hal-01932608

HAL Id: hal-01932608

<https://hal.science/hal-01932608>

Submitted on 23 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quels problèmes pour la recherche d'information médicale au Mali ?

Seydou Doumbia, Lorraine Goeuriot, Marie-Christine Fauvet

LIG (MRIM), University of Grenoble Alpes, 38000, Grenoble, France

ABSTRACT. Bupa company conducted a survey between June and July 2010 in 12 countries around the world. According to this survey, at least three out of five respondents were searching for information about their health on the Internet. Mostly, respondents seek information on a disease or the symptoms of a disease and then on medications. In most cases, a general search engine is used. Another study conducted in the United States in 2013 showed that 75% of respondents who had already consulted the internet for a diagnosis did not feel the need to check the sources. Medical information retrieval is facing specific challenges: the quality and reliability of the results are crucial, and they should take into account the diversity in knowledge (and terminology) of health professionals in front of patients. However, IR application to the medical domain allows to use of terminological and semantic resources (e.g. UMLS) for advanced processing. The aim of this article is to list the main research issues that arise in the context of medical IR on the internet, each of which being illustrated by a usage scenario.

RÉSUMÉ. L'entreprise Bupa a effectué une enquête entre juin et juillet 2010 dans 12 pays à travers le monde. Selon cette enquête au moins trois répondants sur cinq faisaient des recherches d'information relatives à leur santé sur Internet. La plupart de ces personnes cherchent en majorité des informations sur une maladie ou les symptômes d'une maladie puis sur les médicaments. Dans la plupart des cas, un moteur de recherche général est utilisé. Une autre étude menée aux États-Unis en 2013 montre que 75% des interrogés ayant déjà consulté internet pour effectuer un diagnostic n'ont pas ressenti le besoin de vérifier les sources. La recherche d'information dans le domaine médical est confrontée à des enjeux qui lui sont propres : la qualité et la fiabilité des résultats sont cruciales, et ils doivent tenir compte des variations de connaissances (et de terminologies) des professionnels de la santé face aux patients. L'application de la RI au domaine médical permet toutefois, en la restreignant à une langue de spécialité, de disposer de ressources terminologiques et sémantiques (ex : UMLS) permettant des traitements riches. L'objectif de cet article est de lister les principaux problèmes de recherche qui se posent dans le contexte de la RI médicale sur internet. Chaque problème est illustré par un scénario d'usage.

KEYWORDS: Medical Information Retrieval, multilingual IR, use cases

MOTS-CLÉS : Recherche d'Information Médicale, RI multilingue, scénarios d'utilisation

1. Introduction

L'entreprise Bupa a effectué une enquête entre juin et juillet 2010 dans 12 pays à travers le monde (McDaid and Park, 2011). Selon cette enquête au moins trois répondants sur cinq faisaient des recherches d'information relative à leur santé sur Internet. La plupart de ces personnes cherchent en majorité des informations sur une maladie ou les symptômes d'une maladie puis sur les traitements et utilisent généralement des moteurs de recherche généraux. Une étude menée aux Etats-Unis en 2013 montre que 75% des interrogés ayant déjà consulté internet pour effectuer un diagnostic n'ont pas ressenti le besoin de vérifier leurs sources (Fox and Duggan, 2013).

La récurrence de la recherche par les patients d'informations relatives à leur santé en fait un défi majeur de la recherche d'information. Les enjeux liés à de telles recherches sont importants : une erreur (par un professionnel de la santé) sur un diagnostic ou un traitement commise (par un professionnel de la santé) peut avoir des conséquences graves. Il a de plus été montré que la recherche d'information médicale en ligne pouvait générer de l'anxiété chez les patients (White and Horvitz, 2008).

La recherche d'information (RI) dans le domaine médical est confrontée à des enjeux qui lui sont propres : la qualité et la fiabilité des résultats sont cruciales, et les systèmes doivent tenir compte des variations de connaissances (et de terminologies) des professionnels de la santé et des patients. L'application de la RI au domaine médical permet toutefois, en la restreignant à une langue de spécialité, de disposer de ressources terminologiques et sémantiques permettant des traitements riches.

Nous nous intéressons dans cet article aux problématiques posées par la recherche d'information médicale dans un contexte précis : celui du Mali. Certains pays africains comme le Mali sont confrontés à des caractéristiques spécifiques telles que :

- Le manque de ressources : au Mali, les médecins spécialistes ne produisent pas d'informations médicales destinées aux patients locaux (Geissbuhler *et al.*, 2007). On remarque donc un manque de contenus concernant certaines maladies propres au contexte local.

- Le contexte multilingue : au Mali, il existe 12 langues nationales et un alphabet. Certains patients peuvent connaître le nom d'une maladie dans une des langues nationales mais pas en français ou anglais qui restent les langues les plus utilisées sur Internet.

- Le niveau de connaissances médicales : le taux d'alphabétisation au Mali est très faible. Les gens n'ont pas forcément la capacité à lire et comprendre des textes médicaux.

Nous nous intéressons à ces problématiques afin d'étudier comment améliorer la RI médicale au Mali.

Le reste de cet article est construit comme suit : dans la section 2, nous présentons tout d'abord les problématiques identifiées comme étant liées à la recherche d'information au Mali ; dans la section 3, nous illustrons ensuite chacune

de ces problématiques à l'aide de scénarios d'utilisation ; nous terminons finalement par une discussion et une conclusion dans la section 4.

2. Définition du problème

Nous avons identifié au moins trois problématiques liées à la recherche d'information médicale au Mali. Chacune est décrite dans la suite.

2.1. Manque de ressources

D'une manière générale l'Afrique manque cruellement de contenus locaux en ligne. C'est pour cette raison que plusieurs grandes entreprises comme Google ou Orange ont lancé plusieurs projets pour inciter les utilisateurs locaux à créer des contenus pour l'Afrique. En 2010, Google a lancé un moteur de recherche pour les africains appelé *Baraza*. À l'époque, Tidjane Dème, responsable de Google Afrique francophone disait ceci : "Nous avons découvert que les Africains posent chaque jour beaucoup de questions sur Google. Malheureusement ils ne trouvent pas de réponses à cause du manque d'un contenu spécifiquement local ou africain en ligne. Nous sommes convaincus que les internautes locaux sont les mieux habilités pour répondre à certaines questions qui se posent au sein de leur communauté. Et "Baraza" est la meilleure façon de partager et de contribuer à l'information."¹. L'idée était de permettre aux africains de poser des questions et de répondre aux questions des autres. Mais le service a été arrêté en 2012 par manque d'utilisateurs actifs. Il y avait en effet moins de 1000 utilisateurs 20 mois après son lancement².

Après cet échec Google a fait une autre tentative pour pousser les africains à créer des contenus locaux à travers un partenariat avec Orange. En février 2016, Orange a publié un communiqué de presse portant sur la signature d'un partenariat stratégique avec Google³. Les partenaires ont précisé que l'objectif du partenariat était d'abord d'attirer plus de clients africains en proposant des formules (téléphone mobile et forfait de communication) à des tarifs attractifs mais également d'encourager le développement des services et des contenus locaux en étoffant l'offre numérique en termes de contenus locaux et de services en ligne, tels que Google Search, Youtube et Google Maps.

Le domaine de la santé est particulièrement affecté par le manque de contenus locaux. En 2006, la Fondation Health On the Net (HON) a lancé un projet dont l'un des objectifs était de mieux comprendre les enjeux liés à la globalisation des informa-

1. http://www.balancingact-africa.com/news/telecoms_fr/20314/afrique-google-lance-baraza-un-nouveau-site-de-recherche

2. <http://techloy.com/2012/07/10/rip-google-baraza-2010-2012>

3. https://www.orange.com/fr/content/download/35160/1116834/version/2/file/CP_Orange%20%20Google%20partnership%20MEA.pdf

tions de santé en Afrique francophone. Une étude a été publiée dans le cadre de ce projet. Cette étude a montré (1) que l'absence d'un véritable marché de la santé et de l'Internet en Afrique francophone limite le nombre et le type de sites produits localement ; (2) que les médecins spécialistes ne produisent pas d'informations médicales destinées au marché local. En général, ces professionnels s'intéressent aux informations spécialisées produites au nord de l'Afrique et ne produisent pas de contenus à caractère local. Les informations médicales en ligne sont donc produites par les professionnels de la santé du nord et portent sur des sujets qui ne sont pas toujours adaptés au contexte local du sud de l'Afrique.

2.2. Contexte multilingue

Le Mali présente une diversité linguistique très importante illustrée dans la figure 1. Selon le recensement général de la population et de l'habitat de 2011 du Mali, 20 langues sont parlées au Mali. Parmi ces 20 langues, 13 sont reconnues (par le décret numéro 159 PG-RM du 19 juillet 1982) comme langues nationales⁴. Ces langues disposent chacune d'une écriture alphabétique depuis 1967. D'après les données du recensement général de la population et de l'habitat du Mali en 2011, la langue bambara ou bamanankan est la langue la plus parlée au Mali⁵. Le bambara est parlé par près de 59% de la population. Il s'agit de la langue véhiculaire du pays et largement utilisé dans les activités quotidiennes. Il est aussi très présent dans les médias. Il faut cependant noter que le bambara est très influencé par le français, langue officielle du pays. Plusieurs termes français sont utilisés dans des conversations en bambara y compris de termes pour lesquels un correspondant bambara existe. Les charnières de phrases sont également empruntées massivement : "pour que", "comme", "mais", etc. En effet le français, langue officielle du pays, demeure la langue dominante par le fait qu'il s'agit de la langue de l'administration, de la presse et de l'enseignement primaire, secondaire et universitaire. Par contre, on remarque la faible présence des langues étrangères y compris le français, l'arabe et l'anglais (voir Figure 1, autres langues étrangères). Cela est dû au fait que ces langues ne sont parlées au quotidien que par peu de gens en dehors du système scolaire et des médias.

2.3. Niveau de connaissances médicales

La lisibilité et la compréhension des informations disponibles en ligne sur la santé est un enjeu majeur. Le score de Flesch-Kinkaid permet de calculer le niveau de lecture (RGL : Reading Grade Level) en fonction de la longueur moyenne des phrases (ASL, le nombre de mots divisé par le nombre de phrases) et le nombre moyen de

4. <http://http://www.axl.cefan.ulaval.ca/afrique/mali-decret1982.htm>

5. https://www.humanitarianresponse.info/system/files/documents/files/Tableaux_D%C3%A9mographiques_VF.pdf

Population residente de 6 ans et plus selon la langue

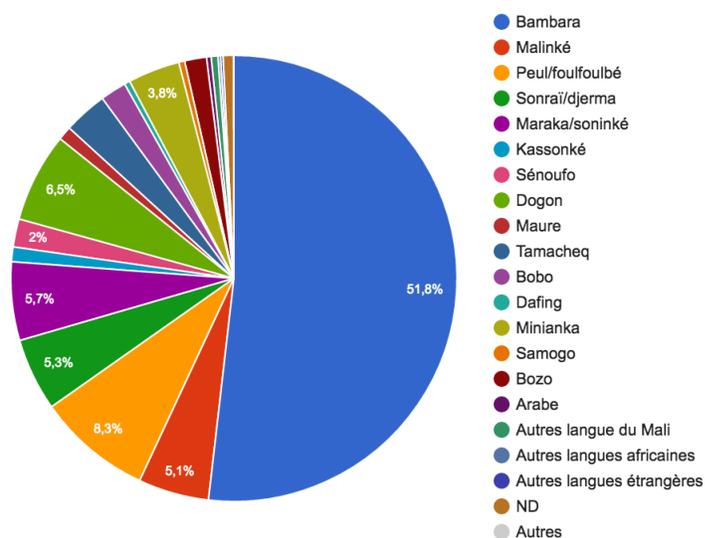


Figure 1 – Répartition de la population selon la langue, au Mali

syllabes par mot (ASW, le nombre de syllabes divisé par le nombre de mots) (Flesh, 1948) :

$$RGL = (0.39 \times ASL) + (11.8 \times ASW) - 15.59 \quad [1]$$

Il a été constaté que la plupart des patients avaient un niveau de lecture moyen de 5 à 6. Or plusieurs études ont montré que le niveau de lecture des documents trouvés sur le web dépassait le niveau de lecture des patients (Overland *et al.*, 1993). Une autre étude a montré qu'un échantillon d'information pour le patient provenant du Web était rédigé à un niveau de 10 (Graber *et al.*, 1999). Dans l'utilisation de sites Web en Irlande, le niveau de lecture des internautes était d'environ 10 (O'Mahony, 1999). Par ailleurs, aucun site de langue anglaise disponible en Espagne n'avait une lisibilité au-dessous du niveau 10 (Berland *et al.*, 2001). Nous n'avons pas trouvé d'études concernant le cas du Mali. Mais en se basant sur les chiffres concernant le taux d'alphabétisation (33,56% en 2011), le taux d'inscription en primaire (77,19% en 2014), secondaire (43,51% en 2014) et universitaire (6,87% en 2012), nous pouvons supposer que le niveau de lecture des patients maliens ne dépasse pas le niveau moyen donné dans les études précédentes.

3. Scénarios d'utilisation

Pour illustrer les trois problématiques définies dans la section 2, nous allons décrire trois scénarios d'utilisation. Pour cela, nous définirons la recherche et son contexte, puis effectuerons la recherche à l'aide de requêtes sur le moteur de recherche Google. Pour chaque recherche effectuée, nous analysons les 100 premiers résultats.

Afin de rendre ces scénarios le plus représentatifs possible, les requêtes sont formulées en bambara (langue parlée par 51.8% de la population malienne, comme détaillé dans la section 2.2). De plus, nous interrogeons la version malienne de Google⁶.

Toutes nos requêtes ont été testées en utilisant des réseaux virtuels au Mali pour que nos résultats ne soient pas influencés par la loi européenne de protection des données.

Bien que nous distinguons ces trois problématiques dans nos scénarios, ces dernières sont étroitement liées.

3.1. Scénario 1 : manque de ressources

Un internaute a souvent entendu parlé autour de lui de la maladie *mara* (onchocercose), il décide d'aller sur internet afin d'obtenir plus d'informations sur cette maladie.

L'onchocercose est une maladie des yeux et de la peau causée par le vers filaire *onchocerca volvulus*. Elle atteint 37 millions de personnes, essentiellement en Afrique, avec quelques foyers en Amérique centrale (Taylor *et al.*, 2010). Environ un demi-million de personnes sont aveugles ou atteintes de troubles de la vision à cause de cette maladie⁷.

Tout d'abord nous décidons d'utiliser comme requête *mara*. Nous obtenons environ 230 millions de résultats. Les 100 premiers résultats montrent que ce terme est très ambigu. Aucun de ces résultats ne parle de la maladie. Le 11e résultat correspond à la page wikitionnaire du mot *mara*. D'après cette page, nous remarquons que le mot *mara* a une signification dans 23 langues y compris le bambara. Par contre la seule définition donnée par wikitionnaire est région ou cercle, qui est une autre définition du mot *mara* en bambara.

Compte tenu de cette ambiguïté, nous décidons de spécialiser notre requête en y ajoutant le mot *bana* qui signifie *maladie* en français. Nous obtenons 493 000 résultats avec la requête *mara bana*. L'analyse des 100 premiers résultats ne nous apprend rien sur la maladie.

La requête composée seulement du mot *mara* nous donne des résultats sur différents sujets tels que le bouddhisme (une page wikipedia qui définit *mara* comme

6. <http://www.google.ml>

7. <http://www.who.int>

mort en bouddhisme), le cinéma (un des résultats parle de la série Angolo-portugaise *mara*), ou encore une personnalité publique (plusieurs résultats parlent de Rooney Mara).

La majeure partie des 100 premiers résultats obtenus avec la requête composée des mots *mara* et *bana*, nous parle de la musique et nous renvoie en général vers des sites internet en langue hindi de téléchargement de musique. Une petite partie des 100 premiers résultats nous parle des articles et films en lien avec l'acteur Rooney Mara.

Malgré la spécialisation de notre requête, nous n'avons trouvé aucun résultat pertinent. Ce scénario illustre bien le manque de ressources et de contenus locaux en ligne.

3.2. Scénario 2 : contexte multilingue

Au Mali le paludisme (*sumaya* en bambara) représente 40% des motifs de consultation dans les structures de santé⁸. Il est la première cause de morbidité et de mortalité chez les femmes enceintes et les enfants de moins de cinq ans. Selon l'annuaire statistique du Mali en 2014, le nombre de cas de paludisme était de plus de 2.5 millions dont 800 000 cas graves⁹. Chez les enfants de un à quatre ans, 801 060 cas ont été enregistrés.

Imaginons maintenant le cas d'un patient qui souhaite avoir des informations sur cette maladie et qui décide d'utiliser la requête *sumaya*. Cette requête nous a renvoyé près de 4 millions de résultats. Nous avons étudié les 100 premiers. Parmi ces résultats les premiers font référence à un restaurant situé à Paris. Les autres abordent de sujets différents sur des personnes dont le prénom est *sumaya*. Seulement 2 résultats sur les 100 premiers font un lien avec le paludisme. Le premier est un article court qui détaille des actions de l'entreprise Novartis contre le paludisme dans plus de 60 pays sur leur site web¹⁰. Le deuxième est un article, dans lequel l'auteur explique la signification réelle du mot *sumaya* à travers une enquête menée dans une région du Mali. L'article ne donne aucune information sur les symptômes ou les traitements de la maladie.

À la requête *paludisme*, traduction française de *sumaya*, nous obtenons près de 7 millions de résultats. Les 100 premiers résultats obtenus donnent des informations sur la maladie. L'analyse des 10 premiers résultats montre qu'ils donnent tous des informations sur les symptômes, la cause, la transmission et la prévention de la maladie.

Dans ce scénario, nous avons commencé par faire une recherche avec le mot en bambara *sumaya* dont la traduction française est *paludisme*, et nous n'avons pas obtenu de résultats pertinents. En traduisant notre requête en français, nous obtenons plusieurs résultats pertinents nous donnant des informations sur les symptômes, la

8. <http://afrique.le360.ma/mali/societe/2016/08/24/>

5082-mali-le-paludisme-represente-encore-40-des-motifs-de-consultations-5082

9. http://http://www.instat-mali.org/contenu/pub/anuair14_pub.pdf

10. <http://www.novartis.com>

cause, la prévention et les traitements de la maladie. Le problème posé ici est que selon la langue utilisée dans la requête le nombre de résultats est très variable (la requête en bambara ne retourne aucun résultat pertinent mais les 100 premiers résultats de notre requête en français sont tous pertinents). Ce scénario illustre que les maladies même très fréquentes ne sont pas documentées en bambara, la langue la plus parlée au Mali.

3.3. Scénario 3 : niveau de connaissances médicales

Imaginons un patient qui n'a pas un bon niveau d'étude, qui parle fréquemment le bambara et qui ressent des douleurs au niveau de son estomac.

Pour ce cas nous choisissons comme requête *fouroudimi foura* ce qui signifie littéralement *médicament contre le mal d'estomac*.



Figure 2 – Résultats obtenus pour la requête *fouroudimi foura*.

En consultant les 2 résultats retournés (voir Figure 2) nous remarquons qu'il s'agit du même rapport de thèse mis à disposition sur 2 sites différents. Ce rapport de 87 pages aborde les risques liés à la vente illicite de médicaments dans les marchés de 2 communes au Mali. Le document est divisé en 6 parties. À la fin de la 1ère partie (l'introduction) l'auteur cite les objectifs de la thèse. Dans la 4e partie l'auteur présente dans un tableau l'identification des médicaments en fonction de leur nom local. Dans ce tableau on trouve la ligne ci-dessous :

Comsil	Fouroudimi foura
--------	------------------

Ce qui signifie que le nom local du médicament (vendu illicitement dans les marchés) *Comsil* est *Fouroudomi foura*.

Selon notre scénario, notre patient n'a pas assez de connaissances pour pouvoir lire et comprendre les résultats de sa recherche afin de se rendre compte si ces résultats répondent ou non à sa requête. Ce scénario est pertinent pour illustrer aussi le manque de ressources.

4. Conclusion

Dans cet article, nous avons décrit trois problématiques rencontrées dans le cadre de la recherche d'information médicale au Mali. Tout d'abord, nous constatons un manque de ressources, en particulier adaptées aux besoins spécifiques des utilisateurs. De plus, la faible production de ressources locales et la multitude de langues parlées au Mali complique l'accès aux informations. Même dans l'une des langues les plus parlées, il peut être difficile d'obtenir des informations. Enfin, on observe au Mali comme partout une forte disparité des connaissances médicales entre les internautes, entre les professionnels de la santé et les patients, mais aussi parmi les patients chez qui le taux d'alphabétisation et le niveau scolaire induit de grandes variations. Les scénarios d'utilisation ont permis d'illustrer par quelques cas simples ces trois problématiques.

5. References

- Berland G., Elliott M., Morales L., Algazy J., Kravitz R., Broder M., Kanouse D., Muñoz J., Puyol J., Lara M., Watkins K., Yang H., McGlynn E., "Health information on the Internet: accessibility, quality, and readability in English and Spanish", *JAMA*, vol. 285, n° 20, p. 2612—2621, 2001.
- Flesh R., "A new readability yardstick", *Journal of Applied Psychology*, n° 32, p. 221-233, 1948.
- Fox S., Duggan M., Health Online, Technical report, Pew research center report, 2013.
- Geissbuhler A., Bagayoko C.-O., Mayor A., Sandru C., Grandbois Y., Roberts I., Droz Y., Steiner B., Weyer F., Gutierrez D., Boyer C., Reffas S., Niang M., Traoré S.-T., Touré Y., Coulibaly B., "Pour des informations médicales en ligne dignes de confiance - Une étude des spécificités en Afrique francophone", 2007.
- Graber M. A., Roller C. M., Kaeble B., "Readability Levels of Patient Education Material on the World Wide Web", 1999.
- McDaid D., Park A., "Online health: Untangling the web. evidence from the bupa health pulse 2010 international healthcare survey", 2011.
- O'Mahony B., "Irish health care Web sites: a review.", *Irish medical journal*, 1999.
- Overland J., Hoskins P., McGill M., Yue D., "Low Literacy: A Problem in Diabetes Education", *Diabetic Medicine*, vol. 10, n° 9, p. 847-850, 1993.
- Taylor M. J., Hoerauf A., Bockarie M., "Lymphatic filariasis and onchocerciasis", *The Lancet*, vol. 376, n° 9747, p. 1175-1185, 2017/02/27, 2010.
- White R., Horvitz E., "Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search", *ACM Transactions on Information Systems*, November, 2008.