Langage communautaire, confiance et recettes de cuisine

Damien Leprovost, Thierry Despeyroux, Yves Lechevallier

Langage communautaire, confiance et recettes de cuisine

Damien Leprovost*, Thierry Despeyroux*
Yves Lechevallier*

*Inria – Rocquencourt, Équipe-projet AxIS, BP 105 – 78153 Le Chesnay Cedex, France
{prénom.nom}@inria.fr

Résumé. De nos jours, les sites de partage de connaissance communautaires représentent une part importante et grandissante du Web. Sur ces sites, les utilisateurs échangent des connaissances, en étant à la fois auteurs et lecteurs du contenu. Dans de telles circonstances, la communauté se structure autour d'une sémantique empirique qui lui est propre, et qui peut différer grandement des standards académiques des domaines concernés. L'analyse de cette sémantique à partir des bases de connaissance de référence traditionnelles peut alors se révéler insuffisamment pertinente pour prendre en compte ces comportements utilisateurs.

Dans cet article, nous présentons une méthode pour construire notre propre compréhension de la sémantique des contributions des utilisateurs, sans recours à une base de connaissance externe. Cette compréhension est rendue possible par une extraction de la connaissance présente dans les contributions analysées. Nous proposons une évaluation de la confiance imputable à cette compréhension déduite, afin d'évaluer la qualité du contenu de l'utilisateur. Ce taux de qualité ainsi calculé peut être considéré comme la mesure avec laquelle le contenu est compréhensible par la globalité des utilisateurs de la communauté. Nous illustrons notre travail en analysant des recettes de cuisine fournies par des utilisateurs sur des sites Web de partage de recettes de cuisine.

1 Introduction

Avec l'essor du Web 2.0, les internautes sont désormais au centre de l'utilisation du Web, étant à la fois consommateurs et contributeurs du contenu du réseau. Au sein de ce Web social, de nombreux sites de partage de connaissance entre utilisateurs, appelés sites de partage de connaissance communautaires, ont vu le jour et gagnent sans cesse en popularité. Le plus célèbre d'entre eux, l'encyclopédie Wikipédia, est même devenue aujourd'hui le sixième site Web le plus consulté du Web mondial [1]. Sur ces sites, les utilisateurs déposent des connaissances, assimilent celles des autres utilisateurs, en discutent et se structurent en communautés. Il est alors observable que ces communautés sont régies par des modes de fonctionnement qui leurs sont propres et issues des habitudes générales des membres de la communauté. Dans ce contexte, les usages deviennent normes, indépendamment d'éventuels standards préexistants.

1. http ://www.alexa.com/topsites

Ce constat s'applique également à la sémantique des échanges, qui dépend donc uniquement des habitudes propres aux utilisateurs impliqués dans l'échange de connaissance. Dans le cadre de la recherche traditionnelle de fouille de la connaissance, cette évolution en relative autonomie peut se révéler problématique. Car dans de telles circonstances, il n'existe aucune garantie que l'évolution des échanges de la communauté se structure autour d'une sémantique qui soit en adéquation avec les bases de connaissance de référence traditionnellement identifiées. La pertinence des conclusions issues de ces mêmes références peut alors ne pas ou peu refléter l'évolution réelle du comportement des utilisateurs et de la sémantique de leurs échanges.

Afin de contourner cette faiblesse des bases de connaissance de référence, nous proposons dans cet article une méthode pour construire notre propre compréhension des contributions des utilisateurs, basée uniquement sur les données de celles-ci. Le fait de s'affranchir des bases de connaissance externes permet de ne plus être soumis aux faiblesses qui y sont liées. En effet, face à un système social en permanente évolution, une base statique de connaissance présente un fort risque d'obsolescence. De plus, nombre d'entre-elles ont bien souvent un caractère ciblé, avec un vocabulaire précis et spécialisé, certes très pertinent mais parfois peu adapté au vocabulaire commun de la majorité des échanges utilisateurs. Notre approche se base uniquement sur le vocabulaire du jeu de données, ce vocabulaire est donc directement issu des utilisateurs. Ces derniers étant à la fois contributeurs et consommateurs de ce vocabulaire, nous visons donc à obtenir la même compréhension que celle que l'utilisateur du système possède lorsqu'il manipule ces données. Afin d'évaluer la qualité du contenu utilisateur analysé, nous évaluons la compréhension que nous obtenons de ces données. Nous attributions pour chaque compréhension déduite une valeur de confiance qui représente dans quelle mesure les utilisateurs de la communauté témoignent d'une utilisation — et donc d'une compréhension — commune de ces éléments. Nous plaçons notre travail dans le contexte des recettes de cuisine, dont les sites de partage communautaires sont nombreux et très populaires sur le Web français comme mondial. Nous illustrons notre travail en analysant des recettes de cuisine fournies par des utilisateurs de ces mêmes sites.

Cet article est organisé comme suit : la section 2 présente un état de l'art de l'utilisation de recettes de cuisine dans le domaine de la gestion des connaissances, puis la section 3 introduit notre modèle d'acquisition de la connaissance. Notre évaluation de la confiance de cette acquisition est décrite dans la section 4, et notre modèle d'expérimentation dans la section 5. La section 6 conclue et présente nos orientations futures.

2 État de l'art

La recette de cuisine est un type de données particulier, composé d'un ensemble d'ingrédients et de procédures d'exécution. Les tentatives de prise en compte des spécificités de ce type de données existent dans la littérature, notamment dans le domaine des systèmes de recommandation. Le *Cooking Assistant* (Sobecki et al., 2006) définit un système de recommandation démographique de recettes de cuisine, basé sur une inférence à logique floue. Raisonnant à partir de métadonnées annotées manuellement, cette méthode est efficace pour fournir une réponse globale à un besoin général. Mais la généralisation des caractéristiques des recettes conduit à une recommandation également généralisée. Il apparaît un besoin de prise en compte des caractéristiques propres aux ingrédients. Freyne et Berkovosky utilisent pour cela dans de multiples travaux (Freyne et Berkovsky, 2010a,b; Berkovsky et Freyne, 2010) la rela-

tion de composition qui existe entre ingrédients et recettes pour propager des évaluations. Par le biais du logiciel d'apprentissage Weka (Hall et al., 2009) et en utilisant l'algorithme d'arbre de décision M5P (Quinlan, 1992) qui y est implémenté, les auteurs déterminent un comportement utilisateur (Freyne et al., 2011). L'ensemble de ces approches nécessitent néanmoins une phase constante de normalisation, un travail d'expert consistant à vérifier ou annoter les ingrédients afin qu'ils correspondent à une liste de référence connue à l'avance.

Bien qu'elles ne soient pas directement liées à nos travaux, il existe également dans le domaine du raisonnement à partir de cas plusieurs approches intéressantes relatives à ce type particulier de données que sont les recettes de cuisines. Le système *CHEF* (Hammond, 1986) est un système d'adaptation par la critique dans le domaine des recettes de cuisine du Sichuan. Cette approche permet notamment de prendre en compte la spécificité du type de données qu'est l'ingrédient, en relevant par exemple les problèmes découlant d'une substitution d'ingrédient, quand bien même ceux-ci aurait été très proches. En revanche, comme nombre de systèmes du genre, une importante phase d'apprentissage est requise. Le système *MIKAS* (Khan et Hoffmann, 2003) pour *Menu construction using Incremental Knowledge Acquisition System*, propose de contourner ce besoin d'apprentissage initial par un recours à l'expert en fonction des besoins d'exploitation tout au long de l'utilisation. Cette aspect de la transmission de connaissance de l'expert au système par l'expérience plutôt que par le déclaratif est vu comme plus efficace et plus adapté à l'utilisation en conditions réelles, plus robuste aux cas inhabituels et imprévus. Il ne permet toutefois pas une évaluation indépendante des contenus, car dépendant des connaissances propres de l'expert.

3 Extraction de l'information semi-structurée

Au sein d'une recette de cuisine, une structure est identifiable : *a minima*, la recette se compose d'un titre, d'une liste d'ingrédients et d'instructions de réalisation. Au sein de ces éléments en revanche, il n'existe pas de structure contrainte. Notamment dans la cas des recettes de cuisine saisies par les utilisateurs, les modes d'expressions peuvent êtres divers et variés. Dans le cadre de nos travaux, nous nous intéressons aux informations relatives aux ingrédients utilisés, à partir des données brutes issues de ces utilisateurs. Ces données sont donc composées de lignes d'ingrédients librement saisies par des auteurs multiples. Pour exploiter ces données, nous recherchons ce que nous définissons comme étant la structure présumée par les utilisateurs. En effet, bien que les formes et les manières de présenter un ingrédient dans une ligne brute sont nombreuses et variées, nous observons toujours un fort consensus sur le mode le plus simple d'expression. Par exemple, s'il est tout à fait possible de trouver un ingrédient décrit comme suit : « un morceau de bœuf d'environ 250g », la majorité des utilisateurs écrivent simplement : « 250g de bœuf ». À partir de ce constat, nous définissons la structure présumée comme étant : *quantité – unité – ingrédient*, où *quantité* est le nombre d'éléments impliqués, *unité* est l'unité de mesure associée à la quantité, et *ingrédient* l'ingrédient lui-même. Pour chaque ligne correspondant au modèle, il est possible de retrouver l'élément vide pour un ou plusieurs des emplacements de structure précédemment défini. Le tableau 1 illustre ce découpage structurel.

Bien sûr, toutes les lignes d'ingrédients ne correspondent pas à ce modèle. Les éléments constituant une ligne peuvent ne pas être dans le même ordre, ou être quantité variable (par exemple, la ligne « 2 ou 3 pommes »). Nous discriminons donc nos lignes d'ingrédient en

TAB. 1 – *Structure présumée*

ligne	quantité	unité	ingrédient
250g de bœuf	250	g	bœuf
3 pommes	3	ø	pommes
pincée de sel	ø	pincée	sel
ketchup	ø	ø	ketchup

quatre classes : les lignes *quantité–unité–ingrédient*, les lignes *quantité–ingrédient*, les lignes *quantificateur–ingrédient* [2] et les lignes non comprises, pour l'instant supposées comme un ingrédient unique non-identifié.

Nous procédons ensuite à une phase d'apprentissage, où toutes les lignes bien formées permettent de comprendre les autres. Cet apprentissage se fait en deux étapes principales :

– Tout d'abord, nous cherchons des incohérences dans les éléments identifiés. La présence d'un ingrédient complexe dans une ligne bien formée permet de mettre en évidence une erreur de détection du même ingrédient dans une ligne plus simple. Par exemple, « 500g de corned beef » identifie clairement « 500 » comme étant la *quantité*, « g » comme étant l'*unité* et « corned beef » comme étant l'*ingrédient*. En revanche, en présence d'une ligne simple « corned beef », basiquement, « corned » sera identifié comme étant un *quantificateur* et « beef » comme étant l'*ingredient*. La connaissance de la ligne complète mentionnée précédemment nous permet alors de comprendre « corned beef » en tant qu'ingrédient dans son ensemble et donc de considérer cette ligne comme étant une *ligne à ingrédient seul*.
– Une fois l'ensemble des incohérences traitées, pour toutes les *lignes à ingrédient seul* restantes, nous cherchons à les faire correspondre aux cas précédemment rencontrés. À partir des *quantités*, *unités* et *ingrédients* précédemment rencontrés, nous les distinguons en *lignes quantité–unité–ingrédient*, *lignes quantité–ingrédient* et *lignes quantificateur–ingrédient*, ou simplement en tant que *lignes à ingrédient seul déduites* si la ligne est simplement un ingrédient seul, dont l'existence a déjà été rencontrée précédemment. À défaut, les lignes restantes sont considérées comme étant des *lignes à ingrédient seul supposées*.

Le tableau 2 présente les classes de lignes ainsi obtenues. Alors que la première étape permet de valider les placements des éléments dans les classes 1, 2 et 3 ; la seconde étape permet de ventiler la classe 5 (qui contient initialement les lignes pour lesquelles aucune compréhension n'est ressortie) dans les autres classes. La figure 1 illustre leur distribution sur le jeu de données Marmiton ainsi que l'effet de la phase d'apprentissage sur le peuplement de ces classes. Dans cet exemple, la taux de lignes non-validées passe de 15% à 1,9% lors de traitement.

4 Évaluation de la confiance

De l'exploitation de la structure présumée précédemment décrite, nous identifions une liste d'ingrédients manipulés par les utilisateurs lors de la rédaction de recettes. Bien que la saisie de ces ingrédients soit entièrement libre, nous observons un effet longue traîne : un petit nombre

2. Une *unité* sans présence de *quantité* est alors appelé *quantificateur*.

TAB. 2 – *Classes de lignes analysées*

Class 1	ligne quantité–unité–ingrédient
Class 2	ligne quantité–ingrédient
Class 3	ligne quantificateur–ingrédient
Class 4	ligne lignes à ingrédient seul déduites
Class 5	ligne lignes à ingrédient seul supposées

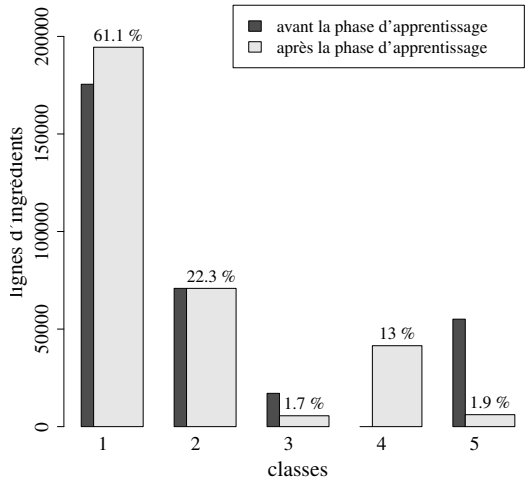

FIG. 1 – *Distribution des classes sur Marmiton*

d'éléments concentre un grand nombre d'occurrences, alors qu'une grande majorité d'éléments ne représente qu'une toute petite partie des utilisations. Ce phénomène est commun à bon nombre de sites sociaux à usages libres — et même au delà — où la distribution suit une loi de puissance. Ce principe de convergence sociale valide notre approche de recherche de motifs majoritaire et guide notre processus d'évaluation de la confiance des ingrédients : une formulation fortement utilisée par l'ensemble des utilisateurs aura nécessairement de bonnes chances d'être bien compris par les utilisateurs et devra donc être gratifiée d'une confiance élevée. À l'inverse un terme extrêmement rare et très peu utilisé ne présente aucune garantie quant à sa compréhension par la communauté et devra recevoir une valeur de confiance faible. La figure 2 illustre la répartition en fréquence cumulée des ingrédients sur le jeu de données de Marmiton. Les valeurs d'*unités*/*quantificateurs* suivent le même schéma de distribution.

Application du principe de Pareto

Eu égard à la distribution en loi de puissance de nos données sociales, nous appliquons à notre modèle le principe de Pareto (ou loi des 80-20), où 80 % des effets sont le produit de 20 % des causes. Appliqué à notre modèle, cela signifie que 80 % des lignes de recettes rédigées

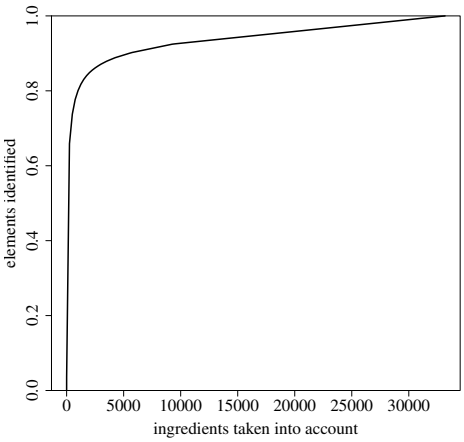

FIG. 2 – *Répartition des ingrédients sur Marmiton*

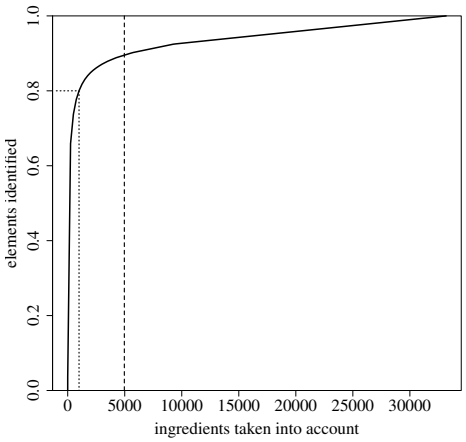

FIG. 3 – *Application du principe de Pareto sur Marmiton*

librement par les utilisateurs sont issus de 20 % de l'ensemble des ingrédients existants. Dans le cas d'une répartition encore plus prononcée (voir figure 2), où moins de 20 % des ingrédients existants représentent plus de 80 % des lignes, cela signifie qu'un certain nombre d'ingrédients ne sont pas significatifs et ne doivent pas être considérés comme tels. Pour discriminer le jeu d'ingrédients suffisamment significatifs à conserver de l'ensemble total des ingrédients renseignés, nous calculons la taille idéale de ce jeu significatif. Nous définissons le nombre a minimal d'ingrédients possible pour représenter 80 % des lignes utilisateurs et définissons ce nombre comme représentant 20 % des ingrédients du jeu significatif. Le jeu significatif est alors l'ensemble d'ingrédient ayant un cardinal b égal à cinq fois ce nombre a qui maximise la représentation des lignes d'ingrédients. La figure 3 illustre cette coupe opérée sur l'ensemble initial.

Confiance par ingrédient Une valeur de confiance est alors attribuée à chaque ingrédient, en fonction de sa fréquence d'utilisation. Cette valeur est maximisée pour le sous-ensemble de tête que sont les 20 % d'ingrédients représentant 80 % des lignes. Cette maximisation est justifiée par la position majoritaire qu'ils occupent dans le système, témoignant d'une utilisation — et donc d'une maîtrise — suffisamment forte par les utilisateurs. Afin de ne pas découper brutalement l'ensemble des ingrédients en deux groupes, le reste des éléments de l'ensemble se voient attribuer une confiance variable en fonction de la fréquence d'utilisation respective de chacun de ses éléments. Pour tout élément i, la valeur de confiance C_i ainsi attribuée est de :

$$C_i = \begin{cases} 1 & \text{si} \quad i < a \\ \frac{N_i - N_b}{N_a - N_b} & \text{si} \quad a < i < b \\ 0 & \text{si} \quad i > b \end{cases} \tag{1}$$

où a est l'ensemble des 20% d'ingrédients les plus utilisés représentant 80% des lignes et N_a le nombre de ses occurrences, b l'ensemble significatif tel que $5 * a = b$ et N_b son nombre d'oc-

curences, et N_i le nombre d'occurrence de i. La confiance est nécessairement contenue dans l'ensemble $[0; 1]$. L'utilisation de valeurs fixes aux extrémités se justifie par la volonté de ne pas sur-nuancer les ingrédients totalement assimilés par la communauté d'une part (confiance à 1), et de ne pas attribuer de valeurs négatives aux ingrédients non-reconnus afin de ne pas sur-impacter le calcul des recettes (confiance à 0). La figure 4 illustre la confiance ainsi calculée des ingrédients de Marmiton.

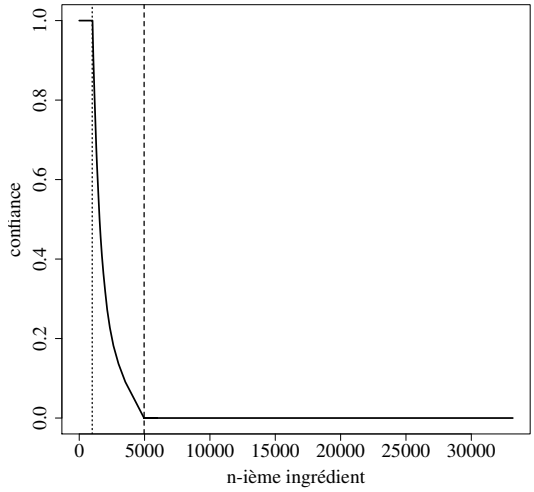

FIG. 4 – *Confiance calculée sur Marmiton*

Confiance par unité/quantificateur Nous considérons à présent l'ensemble des unités — appelés quantificateur en cas d'unité sans quantité — pour leur attribuer également une valeur de confiance. Toutefois, une unité ou quantificateur ne saurait avoir une cohérence universelle et donc une valeur de confiance absolue. En effet et par exemple, si associé à l'ingrédient « lait », l'unité « ml » est fréquente et par conséquent devrait jouir d'une confiance élevée, la déclaration « 200 ml de pommes » n'a aucune cohérence, et donc ne devrait se voir attribuer aucune confiance. Nous identifions donc un lien total et direct entre la cohérence d'une unité ou d'un quantificateur et l'ingrédient avec lequel il est utilisé. Nous considérons donc l'ensemble des unités et quantificateurs en fonction de chaque ingrédient. Pour chaque ingrédient, nous établissons un ensemble de fréquence par quantificateurs, et une valeur de confiance sur le même modèle que pour les ingrédients. Nous remplissons ainsi une matrice de confiance entre ingrédients et unités/quantificateurs. Cette matrice nous permet d'associer à tout couple d'ingrédient x et d'unité/quantificateur i une valeur de confiance $C_{x,i}$, comme étant l'évaluation la cohérence de rencontrer ces deux éléments associés dans une ligne d'ingrédient. Il est important de noter que l'unité/quantificateur vide est répertorié comme toute autre valeur, sa présence pouvait être plus ou moins justifiée en fonction des ingrédients. À titre d'exemple, l'unité vide jouit d'une confiance forte associée à l'ingrédient « poivre » (très rarement lié à

une unité par les utilisateurs), alors que sa confiance est nulle associée à l'ingrédient « riz » (très fréquemment lié à une unité par les utilisateurs).

Confiance par recette À partir des valeurs de confiance des ingrédients et des unités ou quantificateurs par ingrédient précédemment calculées , nous attribuons une valeur de confiance par recette. Cette confiance globale est représentée comme la moyenne des confiances par ligne, qui elle est le produit des confiances de ses composants. La confiance C_x d'une recette x est telle que :

$$C_x = \frac{\sum_{i \in I_x} (C_i * C_{x,i})}{||I_x||}$$

où I_x est l'ensemble des ingrédients de x, C_i la confiance de l'ingrédient i et $C_{x,i}$ la confiance de l'unité associée à l'ingrédient i dans la recette x.

Chaque valeur de confiance ainsi calculée est donc comprise entre 0 (confiance nulle) et 1 (confiance totale). Cette confiance représente donc l'évaluation du degré de certitude qu'un utilisateur du système, auteur comme lecteur, sera en mesure de comprendre et d'utiliser ces ingrédients. La figure 5 présente les valeurs de confiance calculées par recettes sur Marmiton, dont l'expérimentation est détaillée ci-après.

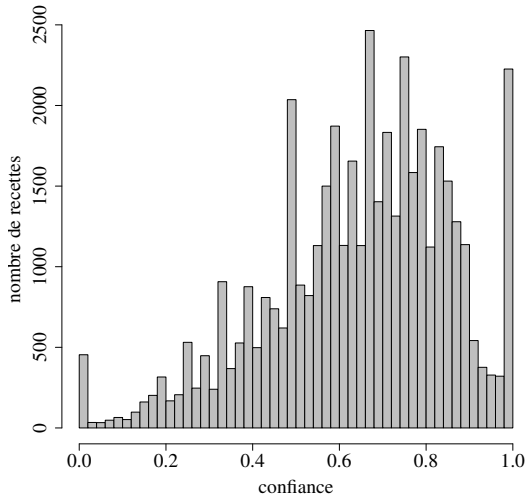

FIG. 5 – *Confiance des recettes du site Marmiton*

5 Expérimentation

Pour valider notre approche et illustrer notre méthode, nous avons réalisé une expérimentation complète sur le site français de partage communautaire de recettes Marmiton.org.

Extraction des données En premier lieu, nous avons développé un robot d'aspiration dédié. Comme il est d'usage dans les sites Web de ce type, chaque page de recette présente un nombre important de liens vers d'autres pages de recettes du site. Nous relevons donc les adresses Web de l'ensemble des pages de recettes présentes sur la page d'accueil du site. Pour chacune d'entre elle, nous indexons le contenu, et relevons l'ensemble des liens vers des pages de recettes que nous ne connaissons pas encore qu'elle contient, et continuons ainsi jusqu'à épuisement de la liste des pages de recettes identifiées. La présence d'éléments de navigation, comme les propositions de menus, les dernières recettes publiées ou les recettes au hasard nous permettent de ne jamais risquer de s'enfermer dans une partie plus réduite du site. En outre, cette approche est adaptée à la philosophie du site, qui veut qu'un maximum de recettes soit visible en un minimum d'effort par ses utilisateurs. Par cette méthode, nous avons collecté 44 169 recettes distinctes.

Normalisation des données extraites Dans un second temps, nous avons utilisé un analyseur syntaxique développé spécifiquement pour parcourir et normaliser l'ensemble des recettes collectées. Cette normalisation au format XML nous permet de structurer les différents types de données que contiennent les recettes, et notamment les lignes d'ingrédients. Par ce traitement, nous avons identifié 354 856 lignes d'ingrédients. Après traitements et corrections d'erreurs mineures, nous avons relevé 33 177 ingrédients distincts dans ces 354 856 lignes.

Calcul de la confiance et résultats Nous avons alors appliqué notre méthode comme présentée dans les sections 3 et 4. La figure 5 présente la distribution des valeurs de confiance par recette calculées sur le jeu de données. Cette distribution suit globalement une loi normale, avec des pics aux extrémités, conséquence des recettes très populaires et très simples d'une part (comme faire une pâte) et incomprises d'autre part [3]. La table 3 résume les principales mesures de cet ensemble des valeurs de confiance des recettes.

TAB. 3 – *Confiances des recettes sur Marmiton*

mesure	**valeur**
premier quartile	0,800
médiane	0,667
troisième quartile	0,507
moyenne	0,658

Outre les résultats présentés tout au long de l'article, nous avons également mené en fin d'expérimentation une première exploration de croisement de ces résultats avec les divers métadonnées additionnelles que propose Marmiton. La figure 6 présente l'amplitude de confiance des recettes en fonction de leur type déclaré.

On remarque une confiance nettement supérieure des desserts, cohérent avec le fait que nombre d'entre-eux partagent des ingrédients très communs, limitant ainsi les apports ésotériques qui grèvent la compréhension, et donc la confiance. Le phénomène inverse s'observe pour les boissons, où la base d'éléments communs est beaucoup plus réduite et où en conséquence une beaucoup plus forte diversité d'ingrédients s'exprime — ingrédients que l'on ne

3. Il s'agit notamment de recettes mal rédigées pour lesquelles l'analyse des ingrédients ne fournit aucun résultat

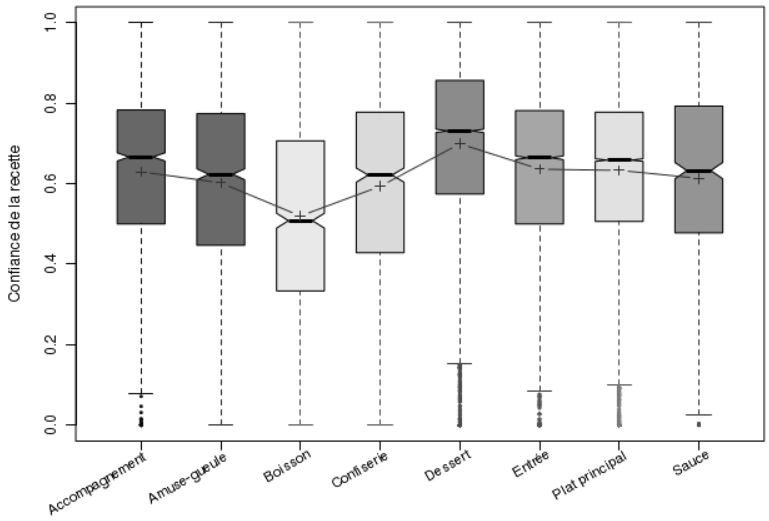

FIG. 6 – *Amplitudes de confiance des recettes par type sur Marmiton*

retrouvera bien souvent pas dans les autres recettes du site — ce qui conduit globalement à une confiance plus faible pour les recettes de ce type.

6 Conclusion et travaux futurs

Dans ce papier, nous avons présenté une méthode pour évaluer la confiance attribuable à une publication utilisateur, comme étant la probabilité qu'un autre utilisateur du système comprenne la sémantique de sa contribution. Nous utilisons une approche indépendante de toute base de connaissance externe, afin de raisonner directement sur les termes manipulés par les utilisateurs. Cette méthode présente également l'avantage de ne pas être dépendant de la langue, ni de souffrir des problèmes de pertinence ou de couverture relatifs aux bases de connaissance de référence, tout en construisant une base de connaissance propre à la communauté.

Pour nos travaux futurs, nous projetons d'exporter la connaissance extraite de cette compréhension des contributions utilisateurs, ce qui permettra de définir sans apport extérieur l'ontologie du système analysé, ou d'enrichir une base extérieure pour améliorer sa pertinence et lutter contre son obsolescence. Enfin, l'application de méthode de partitionnement de fouille de données, guidées par nos mesures de confiance, permettra prochainement d'évaluer une structure interne de la sémantique du système et des relations déductibles qui existent entre les différents ingrédients (proches ou dérivés) ou recettes (variantes, alternatives).

Références

Berkovsky, S. et J. Freyne (2010). Group-based recipe recommendations : analysis of data aggregation strategies. In *Proceedings of the 2010 ACM Conference on Recommender Systems*, Barcelona, Spain, pp. 111–118. ACM.

Freyne, J. et S. Berkovsky (2010a). Intelligent food planning : personalized recipe recommendation. In *Proceedings of the 2010 International Conference on Intelligent User Interfaces*, Hong Kong, China, pp. 321–324. ACM.

Freyne, J. et S. Berkovsky (2010b). Recommending food : Reasoning on recipes and ingredients. In *User Modeling, Adaptation, and Personalization, 18th International Conference*, Big Island, HI, USA, pp. 381–386. Springer.

Freyne, J., S. Berkovsky, et G. Smith (2011). Recipe recommendation : Accuracy and reasoning. In *User Modeling, Adaptation, and Personalization, 19th International Conference*, Girona, Spain, pp. 99–110. Springer.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, et I. H. Witten (2009). The weka data mining software : an update. *SIGKDD Explor. Newsl. 11*(1), 10–18.

Hammond, K. J. (1986). Chef : A model of case-based planning. In *AAAI*, pp. 267–271.

Khan, A. S. et A. Hoffmann (2003). Building a case-based diet recommendation system without a knowledge engineer. *Artificial Intelligence in Medicine 27*(2), 155–179.

Quinlan, R. J. (1992). Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence*, Singapore, pp. 343–348. World Scientific.

Sobecki, J., E. Babiak, et M. Slanina (2006). Application of hybrid recommendation in web-based cooking assistant. In *Knowledge-Based Intelligent Information and Engineering Systems, 10th International Conference, Proceedings, Part III*, pp. 797–804. Springer.

Summary

Today, websites sharing community knowledge are an important and growing part of the Web. On these sites, users share knowledge, being both authors and readers of the content. In such circumstances, the community is structured around an empirical semantics of its own, and may differ greatly from the academic standards of the areas concerned. The analysis of the semantic knowledge bases from traditional reference can then be insufficiently relevant to take into account the user behavior.

In this paper, we present a method to build our own understanding of the semantics of user contributions, without the use of any external knowledge base. This understanding is performed thanks to the knowledge extracted from the same analyzed user contributions. We propose in this method an evaluation of the trust attributable to the deduced understanding, in order to evaluate the quality of user content. This computed quality rate can be viewed as the extent to which the content is understandable by the community of users. We illustrate our work by focusing on the cooking recipes provided by users on sharing websites.