eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

CrossMark

ORIGINAL ARTICLE

HOMER: A semantically enhanced knowledge management approach in the domain of homemade explosives intelligence

Dimitrios Pappas[1] · **Iraklis Paraskakis**[1]

Abstract This paper presents a new approach, in handling data (encoding, managing and retrieving) in secure sensitive and classified organisations (such as Law Enforcement Agencies (LEAs)), that utilises Web 3.0 technologies as well as knowledge management techniques and pushing of information. This approach signals a departure from current use of databases and pulling of information technologies as well as allowing separation of concerns between how data are organised/structured and how data are manipulated/processed. Such an approach utilises an adaptive knowledge management platform capable of supporting organisational operations of LEAs using data aggregated from assorted, heterogeneous and online sources. Such knowledge is then pushed to the users, using recommenders, in an effortless manner addressing the needs of the organisation. Moreover, the system is designed to afford easier change of operational needs through the addition and removal of multiple folksonomies (representing changes in focus or new trends). These changes are further enriched with semantics providing specialised domain-specific content recommendations and semantically enriched search capabilities. This approach to knowledge retrieval has been applied to the domain of homemade explosives and counter-terrorism efforts as part of the HOMER project, where data are aggregated from sources such as police databases, online forums and explosives wikis. Data are stored in an unstructured manner and annotated by the users, ultimately being categorised as per the knowledge retrieval needs of the organisation, which in this case is to carry out efficient and effective investigations regarding homemade explosives. We describe the architecture of a system that can efficiently and effectively support related investigatory activities, and we also present an evaluation from the perspective of the end-users.

Keywords Knowledge retrieval · Recommender systems · Semantic enrichment · Folksonomies · Personalisation · Homemade explosives · Classified organisations

1 Introduction

The process of knowledge management (KM) revolves around providing efficient and effective means of recording, cultivating, sharing—internally or externally—and capitalising on organisational knowledge. Knowledge has been traditionally kept in database systems in a structured manner, which were offered the best and most powerful choice of KM back in the advent of Information Technology (IT) for organisations, and hence, organisational knowledge for companies over the years has been stored in this manner.

Yet, with the ever-growing amount of online resources today and the semantic web, a common framework allowing for such data to be shared with and reused in several different systems, organisations have the potential to benefit from incorporating select such pieces of knowledge in their own pool. Nonetheless, this machine-understood wealth of knowledge is represented in assorted formats, possibly even providing content in an unstructured

✉ Iraklis Paraskakis
iparaskakis@seerc.org

Dimitrios Pappas
dimpappas@seerc.org

[1] South East European Research Centre (SEERC), The University of Sheffield International Faculty, CITY College, Thessaloniki, Greece

manner, and this naturally constitutes an immense detriment to incorporating such information in existing, structured organisational knowledge. This is attributed to the need for changing all that information, as it can be associated with a lot of effort in pre- and post-processing of such information to adhere to the current schema, so that the current system can interlink it properly with the rest of the resources.

Law Enforcement Agencies (LEAs), as an example of such organisations, possess a wealth of knowledge recorded in traditional database systems following a specific schema to represent, categorise and retrieve it. Such systems tend to be inflexible with regard to the type of information they can record and to any emerging needs over the years, as heavy schema modification and re-factoring of captured data are required. Meanwhile, the addition knowledge outside the confines of the organisation itself (e.g. content discovered online) can be a lengthy and difficult process with the lack of adaption such systems suffer from.

Non-database, unstructured approaches to KM contribute to such organisational demands because it relies on the creation of a flexible knowledgebase and such systems can provide efficient and effective means of recording, cultivating, sharing—within the organisation and with external parties—and utilising such knowledge. Covering gaps in personnel expertise, for example, is promoted through the organisational learning capabilities of KM systems, an example for LEAs being that of bolstering expertise across all employees (e.g. first responders) in the organisation instead of the few select ones (e.g. tactical search teams) with extensive knowledge in homemade explosives (HMEs) and related cases.

In a world with an ever-growing, immense amount of information available online, it is necessary for organisations to seek KM solutions that are capable of interlinking their own existing, private knowledge with the ones available online, as effortlessly and efficiently as possible. Semantic techniques that can describe such heterogeneous pieces of information are essential, as is the personalisation and adaptation of such systems to the needs of the end-users; abundance of information can become an impediment to the success of the system if it cannot be accessible by users as and when needed. Consequently, efforts in the development of KM systems ought to steer towards bolstering the capacity for accommodating existing organisational knowledge, incorporate external (i.e. online) knowledge effortlessly and push relevant information to users in a context-aware manner through their various interactions with the platform, whether it is passively through browsing or actively through searching.

The proposed solution from this architecture, implemented as the HOMER KMP to support operations in the homemade explosives domain, follows this approach. The rest of the paper is organised as follows: a discussion on the issues pertaining to the KM systems and the mechanisms for achieving the aforementioned objectives, presentation of the adaptive knowledge retrieval approach, the semantic enrichment of folksonomies and recommender systems that make it work, its implementation for the specialised homemade explosives solution and finally the evaluation of the solution through the implemented system.

2 Background issues

KM efforts are being motivated by the potential to leverage organisational learning, raise expertise across organisation employees, increase network connectivity and foster innovation, among numerous other benefits. Knowledge retrieval, an integral part of the KM process, seeks to provide access to information in a structured form similar to the human mental process of knowledge acquisition and understanding. Unlike the very elementary processes of data and information retrieval, knowledge retrieval differentiates because it is heavily focused on the need to achieve proper representation of knowledge, as well as its accurate extraction from the knowledgebase as per the specific needs of end-users. Therefore, knowledge retrieval aspires to move away from the deterministic or statistical models employed in most traditional systems, artificial or natural language-only queries and representation by numbers, rules and markup language. Consequently, knowledge retrieval is based on semantic and inference models, accessing information via natural language input complemented by knowledge structure and representation, using a variety of more complex mechanisms such as ontologies, semantic networks, predicate logic and concept graphs (Yao et al. 2007).

Users tend to be presented with an information overload when they attempt to find information through the knowledge retrieval tools of their organisation; hence, the retrieval process is as successful as its capacity for producing relevant results with minimal effort. KM techniques focus on producing relevant results to the users to tackle issues with large bulk of information returned from searches. Consequently, any organisational KM system needs to take under consideration all the needs of the users and, by extent, the objectives of the organisation. This process is referred to as personalisation, and it is becoming essential to the successful knowledge retrieval efforts (Bennett et al. 2015). These methods have a vast array of techniques that can be used to achieve them, and they are usually combined for the best result. Some of them include the analysis of user behaviour on the system (e.g. what they have searched for, what they frequently access), as well as user-

related context (e.g. user profile details, their current location), or exploitation of methods used to represent the knowledge on the platform (e.g. semantics, ontologies, taxonomies) (Loew et al. 2007). Essentially, systems designed with such principles in their core are capable of adapting to the users and their needs, and hence solve knowledge navigation problems users may encounter within the knowledgebase of their organisation.

One of the means that KM provides for users towards retrieving only information important or relevant to them specifically is the utilisation of recommender systems. When it comes to pushing relevant information to users, recommendations are the predominant knowledge retrieval bolstering tools employed nowadays, which exploit all aforementioned techniques to suggest more relevant resources to the users during their interactions with the platform (e.g. searching for something, browsing content) (Steichen et al. 2012; Sharma and Singh 2016). Recommender systems tend to fall into one of the following three categories based on which type of information available to the KM system they operate on: (a) content-based filtering, (b) collaborative filtering and (c) hybrids combining elements of both (a) and (b) (Melville and Sindhwani 2010).

Content-based filtering studies the core characteristics of a resource and compares it to those of other resources in order to recommend items that possess similar properties. On the other end of the spectrum, collaborative filtering is based on the study of user behaviour and interactions with the system, in an attempt to predict resources that users would be interested in; this can be investigated in its explicit form (e.g. search, rating, favourites) or implicit form (e.g. resources accessed, preferences of user social network).

Personalising knowledge retrieval through the incorporation of recommender systems has been highly successful during the last years, especially with the advent of online stores and social networking platforms (Steichen et al. 2012). The focus on work carried out in this area is based on achieving better personalisation by fine-tuning analysis and prediction algorithms to raise result precision and recall, as well as the development of hybrid systems that manage to tackle limitations imposed by separate content-based or collaborative filtering systems (Loew et al. 2007; Rajeswari and Hariharan 2016). Annotating content with tags, drawn from a folksonomy (i.e. tag taxonomies that can be modified by users) incorporated in the system, is part of the collaborative filtering process and has also proved to be a popular and useful means of categorising and retrieving resources in various content-heavy web applications.

The problem with several attempts in this field is that they tend to disregard the important factor of the inconsistency of human information retrieval behaviour, as well

as the relation it has to the domain of the organisation and the end-user (Iqbal et al. 2016). Semantics, which refers to the way of providing as close to the real-world description as possible for a term, is one technique employed in KM systems so that knowledge retrieval can be enhanced. They are used extensively to develop links between the properties and meanings behind terms and interlink resources in a knowledgebase. These resources can be totally heterogeneous in nature, such as a police report in structured format and a wiki post about chemical information, or a forum post regarding HME preparation activities and new recipes emerging in the dark web.

Such information has to be integrated as part of the knowledge in LEA KM systems to bolster counter-terrorism capabilities, yet the traditional approaches fall short due to the need of recording information with an inflexible schema-based format that follows a set of very specific rules in order to homogenise everything. Typical semantic web solutions rely on languages such as Resource Description Framework (RDF), Web Ontology Language (OWL) and Extensible Markup Language (XML), by which the documents, as well as the various relationships among them, can be described effectively. Information is thus associated with the machine-readable descriptions of bits and pieces of information scattered around the web, something that developers of KM systems may capitalise on to interlink their own data with that available online. Assorted methods of enriching content with semantics can be combined to have a more effective representation of the knowledge within the chosen domain.

Collaborating filtering in the form of annotating content with terms from a taxonomy is another, user-centric way of interlinking heterogeneous resources within a knowledgebase. Further enriching these terms with semantic properties, therefore giving them meaning within the chosen domain and describing their relationships, can effectively interlink information annotated with these terms by formulating semantic maps in a layer below that of the information itself. This will be discussed in more detail in the following section, where we discuss how focusing on enhancing collaborative filtering with semantics can result in the better representation and retrieval of a domain. The obvious benefit of the collaborative filtering approach in this case is that it is not entirely affected by the content of the document, and it is irrespective of its structure, too. The terms can be used to annotate a free-text document, a police report presented in tabular form, as well as a forum post, or even multimedia resources.

In the world of LEAs and related stakeholders in the domain of HMEs, it is essential to develop a KM system, usually referred to as a knowledge management platform (KMP), which is personalised towards offering more accurate and relevant resources for HMEs investigations to

end-users. Considering all these benefits of recommender systems, collaborative filtering and semantics, we attempt to apply the adaptive development process to this intricate domain and evaluate its capability for supporting efficiently and effectively organisational operations. The adaptive knowledge management approach presented here is this combination of semantics and tag folksonomies, exploited by recommender systems that also consider user behaviour analysis and context to provide better results. The two key mechanisms that utilise the results of these techniques to offer expert knowledge are the semantic search and the recommender systems.

A KMP based on the combination of folksonomies, semantic networks and user behaviour analysis has the potential to be parametrised as needed to meet organisational needs. It offers intuitive and multi-faceted knowledge retrieval that manages to adapt to the domain of the organisation and aspires to tackle the problems of user inconsistency and result precision. This is achieved primarily with the utilisation of semantic search and content recommendations during searching and browsing. The goal is to push relevant information to users, who, with minimal effort, can reach the knowledge they require within the system. The systems employed (tag extraction, natural language processing, folksonomy semantic enrichment, hybrid recommenders) all work in tandem so that we implement a user-centric—at the same time user-friendly—approach to KM. Furthermore, knowledge itself is formed from assorted heterogeneous sources, which are not stored in a structured, database-like manner, but ultimately become interlinked through collaborative filtering and semantic enrichment mechanisms.

Consequently, the three major steps to follow for this approach are:

1. To implement different folksonomies specific to the domain (of HMEs in the case at hand).
2. To protect the folksonomies and enrich them with semantics.
3. To enhance information retrieval with semantically relevant results and recommendations.

The capacity for adaptation is showcased with the application of this approach to the HMEs domain as part of the Homemade Explosives Characterisation and Capability (HOMER) research project. The objective of assisting counter-terrorism efforts (i.e. investigations, threat response, threat prevention) is an integral part of the HOMER project. In the following sections, we will discuss the specifics of the architecture of a KM system following this three-step approach, through the conceptual architecture, folksonomy-based knowledge management, semantics and recommendations. We also present how this system is tailored to the domain of HMEs, becoming

the HOMER KMP, by visiting each of the four aspects of KM (recording, developing, retrieving and sharing knowledge) and each step of the process. We conclude with the evaluation of the system by investigators and chemical experts who are the intended end-users of such a platform in LEAs and other related stakeholders in the field of HMEs.

3 Adaptive knowledge retrieval

3.1 Conceptual architecture

The conceptual architecture of the proposed KMP in Fig. 1 above presents the core of the system, the knowledge base, which enables the smart processing of incoming information through the mechanisms of indexing (with the text analysis, tags, taxonomies components), and which is then utilised by the knowledge retrieval mechanisms and the recommender systems provided to the users (the semantics, browsing history, tags, browsing content, semantic search components).

The users of the system are allowed limited access to the system, its functions and resources, which is governed by their role in their organisation (the role-based access component). Employees of an organisation are assigned specific roles that come with varied rights and responsibilities. This particular mechanism ensures that the KMP can leverage the actualisation of an organisational structure: only certain users may create knowledge, and users may selectively consume knowledge as the organisational policies allow for it.

Furthermore, users of the system naturally take active part in the recording of knowledge that is performed via the content generation components (the blogs, Wikis at conceptual level). Intuitive web technologies for capturing, sharing and developing knowledge, such as Blogs and Wikis, enable the option of categorically organising resources (e.g. exports reports blog, imports reports blog), as well as structuring the information hierarchically if needed (e.g. department descriptions Wiki). Each entry is in the form of a typical document enriched with file attachments and supporting rich-text and HTML content, with the additional capability of directly referencing other related documents if the author deems it necessary. This enables the system to record virtually any type of data source and, eventually through the collaborative filtering and semantic technologies, interlink them and generate a strong, extensive knowledge-base for the desired domain. Information such as YouTube videos, online forum posts, dark web pages discovered through focused crawling—all of them can be part of the recorded knowledge.
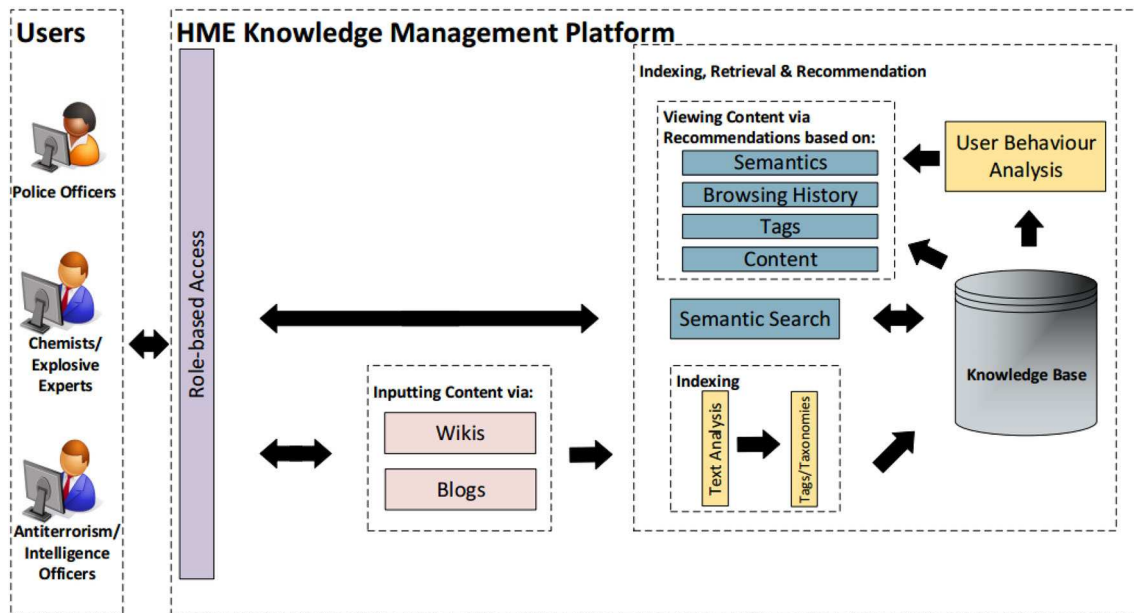
Fig. 1 Conceptual architecture of the proposed KMP

All these components together realise the four important aspects of KM: creating, developing, sharing and using organisational knowledge.

3.2 Folksonomy-based knowledge management

Collaborative filtering utilising tags is a staple of modern web-based information systems, and as such, it has been incorporated in the proposed KMP. As a step of the content generation process in the system, the end-users are required to go through the process of annotating their documents with tags that can best categorise them. In several KM systems, there is a centralised, locked taxonomy that can be used to draw terms from and perform this process. In the proposed system, the approach differs in that the tags are based solely on the utilisation of folksonomies. This, naturally, enables users to have substantial control over the organisation of information in the system, an immense benefit for the purposes of personalising knowledge retrieval or for facilitating adaptivity. This is the result of a tagging process that encourages users to think about the domain itself, how they and other users would have to go about accessing the information they are recording. It thus constitutes a way of reflecting end-user way of thinking about the domain and which is crucial to account for problems.

Collaborative tagging systems based on folksonomies are adaptive when it comes down to organising information and could perform better than orderly classification systems (Halpin et al. 2007). Unlike the typical database or schema-based approaches, the relations and categorisation can evolve and adapt to new data, hence overcoming the problems of having to redesign the system from scratch. Another inherent property of such systems is the option to investigate the frequencies of the tags being used and derive valuable statistics about the domain, the knowledge within the system and emerging trends (Peters 2009). This can be extremely valuable to specific organisations focused on fostering innovation within their confines and for adapting to new trends so that they are not left behind by their competition, as well as to spot emerging markets.

While tagging solutions have been praised for several aspects related to knowledge management, they have also received considerable criticism regarding their effectiveness. Researchers primarily attribute these issues to the potential ambiguity of tag meanings and the possibility of overlapping tags due to synonyms, as well as to the lack of a centralised, controlled vocabulary (Halpin et al. 2007). These arguments state that when such issues take root within a folksonomy, it might be impossible for the users to develop a consensus and for a common vocabulary to emerge. Therefore, knowledge organisation and effective retrieval of that knowledge can be crippled as there is no way to consolidate and categorise information properly. However, these problems can be addressed through several mechanisms and functionalities that can be incorporated in the system.

It can be argued that language-related problems are a by-product of an extremely large and diverse user base on a general-purpose application, whereas domain-specific applications expected at organisational level are bound to follow a much stricter and narrow vocabulary that users

may employ. Domain and content-specific tag recommendations provided to users during the process of content creation also contribute to alleviating such issues (Halpin et al. 2007). Furthermore, empirical studies have refuted the issue of consensus by indicating that, over time and with increasing numbers of recorded resources, folksonomies tend to turn into stable distributions despite the absence of a centralised and controlled vocabulary; in some cases, common structures at a categorisation level can also be developed (Robu et al. 2009).

The benefits of folksonomies to knowledge retrieval on an organisational level prompted us to incorporate this solution as the basic tool for knowledge management on the KMP, while taking under consideration the potential pitfalls it is bundled with. To this end, we have employed several prevention and counteracting mechanisms that possess the potential to convert the folksonomy approach into a robust solution for knowledge organisation and development, in addition to retrieval. As part of the second step and regarding the protection of folksonomies, we will present these counter-measures below.

3.2.1 Multiple, domain-specific Folksonomies

By studying the domain and breaking it down to a condensed set of sub-domains, we are able to narrow down the vocabulary that users may employ to tag resources for each one. Naturally, this enables other KM methods we employ (semantic enrichment, recommendation systems) to capitalise on the variety of vocabularies being formed and produce fine-grained knowledge retrieval processes, as well as to achieve multidimensional knowledge representation. As an example, in the domain of HMEs, a user could have a different set of recommendations or semantic relations in searches only for chemical components involved in terrorist incidents. This is hence a capability of the system to be future-proof since if new requirements emerge for expanding the knowledge on a domain, then it is as simple as initiating a new folksonomy that users can employ to annotate uploaded content with relevant terms. Organising or retrieving the new resources as required without the need of redesigning the whole domain representation and relationships is the core benefit. This also refers to the first step of our adaptive knowledge retrieval approach, where multiple folksonomies are created for the domain of the organisation.

3.2.2 Folksonomy-based tag recommendations

Introducing multiple folksonomies that users are obliged to employ during content generation, provided that the content they are generating applies to all of them, may inadvertently discourage them from utilising the knowledge recording process to its fullest extent. Additionally, there may be the

case where the author misses semantically important terms while annotating the content, therefore additional opportunities of interlinking information and expanding knowledge in the system. In the proposed KMP, we aspire to counter such problems that with tag recommendations based on a hybrid recommender system using a combination of semantics and Natural Language Processing (NLP) techniques. The content is being analysed by the NLP in relation to folksonomy terms and semantics, as well as to existing content residing within the knowledgebase, therefore providing a list of the most important terms to the user. Another welcome by-product of this approach is that it has the added benefit of tackling issues regarding confusing tag meanings and synonyms, which can otherwise usually be a problem in folksonomies by prohibiting the development of a consensus on important terms to describe information.

3.2.3 Content validation

Erroneous or inappropriate tagging by a single user may contaminate the folksonomy and, as a consequence, the semantics and recommender systems. This is also one of the reasons why taxonomies for sensitive, or high-risk domains, are expected to be locked and predetermined only by domain experts, which eventually leads to the lack of adaptivity in such a collaborative filtering approach. Indeed, such problems could result in contamination of the overall knowledge residing in the platform; the outcome could be in the range of insignificant to disastrous, depending on the domain and applications of the system. In order to minimise such risks, a content validation mechanism has been implemented. This policy dictates that expert users review newly created content, essentially all content being uploaded to the KMP, and either approve it or request changes from the author. This does not apply only to the content itself, but also for the tags used for each domain-specific folksonomy. The resource is still accessible by all users with access rights, and they are warned of its validations status, residing in a limbo state in the system; it is not incorporated into existing knowledge unless it has been validated, which essentially prohibits the emergence of this information as a recommended resource or related to the search term(s) from the semantic search facilities of the system.

3.2.4 Folksonomy refinement

As part of the semantic enrichment of the folksonomies in this proposed approach, one technique that has been employed is the refinement of the folksonomy structure. This mechanism essentially attempts to give a structure to the folksonomy, as the consensus on the vocabulary is getting stronger by consulting an external knowledgebase. The end result is a proper hierarchical categorisation of

resources annotated with the specified terms, which in turn enhances the information retrieval capabilities of a system. Recommender systems can also exploit such structures if developed in this manner, and thus provide much more accurate and relevant recommendations to end-users during their interactions with the KMP. Semantic search results also provide better filtering capabilities.

4 Semantic enrichment of Folksonomies

Flat, unstructured and unsupervised vocabularies in folksonomy-based KM systems have been found lacking in result precision due to the lack of hierarchical organisation of knowledge and contamination with ambiguous or irrelevant tags (Jabeen et al. 2016). One way of combating that problem is through the employment of mechanics that can give a structure to the folksonomies. This endeavour is a popular practice within such knowledge management systems, and on the way to realising the semantic web, because of the core benefits that semantics have been proven to bring to folksonomy systems: structure and its protection, as well as search enrichment, all of which indirectly bring higher search and navigation precision (Jabeen et al. 2016). We have discussed some of these counter-measures as they have been employed in the proposed KMP, and we will present how this enrichment has been designed and implemented.

Semantics broadly refers to the study of meaning in language; in the case of tags and the field of knowledge management, it is the study of the relationship between the word and its denotation. By clarifying this relationship of a tag and its meaning within the current context, identifying a concept, as well as tag interrelationships, structure and hierarchy may emerge. Since all that information is exploited by the search and recommendation mechanisms, it is natural that result precision as well as knowledge navigation and representation is ultimately enhanced.

The typical techniques employed in such cases attempt to formulate semantic models, such as semantic networks, to ultimately generate a conceptual view of the data that is a close representation of the real world. The relations identified among concepts vary depending on the objective of the methods followed, usually focusing on finding equivalent terms, subsuming relationships, hierarchical higher and lower-level concepts, to name a few (Jabeen et al. 2016). When developing a KMP and the recommender or semantic search functionalities, the design can account for all aforementioned aspects to provide fine-grained search filtering or highlight otherwise very obscure connections between terms and, by consequence, pieces of knowledge.

Folksonomy enrichment with semantics tends to rely on three major contributing sources (Jabeen et al. 2016):

1. Folks (i.e. the users, their process and motivations).
2. External knowledge-based sources (e.g. DBpedia and Wikipedia, Ontologies).
3. Statistical and mathematical techniques.

The proposed KM system includes components in its architecture, as outlined in Fig. 1, that exploit all of these sources. To begin with, the investigation of sub-domains for the creation of multiple domain-specific folksonomies, in conjunction with the expected role-based access of the users, focuses on the "Folks" source. Understanding the expected user behaviour and their actions within organisational context is the first and most invaluable step. Ultimately, by understanding how users respond to tagging and by observing how they access information, the proper partitioning of the domain emerges. Once the new folksonomies are populated with terms and subsequently enriched, the resulting recommendation and search precision increase becomes apparent through the underlying semantic network formed for each one. Users are also in control of validating resources, expanding the vocabulary, forming a consensus and shaping the structure of their organisational knowledge. This provides a high degree of adaptivity to the retrieval of this knowledge, as users decide how they would like to access their resources over time via the collaborative filtering process.

External knowledge-based sources, specifically DBpedia,[1] as well as statistical techniques, are utilised by both the tag recommendation component and the text analysis indexing mechanism. Tag recommendation uses a tag extraction function based on statistical NLP techniques. It attempts to identify key terms within the provided content and consults DBpedia, using the concepts category and upper or sub-categories in the external source that corresponds to the sub-domain (e.g. DBpedia category: Countries and Cities, folksonomy and sub-domain "locations"). The algorithm assigns a rating to each identified term by comparing the semantic properties of each term to that of the other term, as retrieved through DBpedia. It then returns a list of all relationships between with those possessing high ratings, therefore assisting the users with the process of tagging while protecting taxonomy structure, too.

The text analysis indexing mechanism is the most complex sub-system and leverages the folksonomy refinement technique. Essentially, it performs similar functions as the tag extraction but for different purposes. The NLP component compares knowledge artefacts with each other in order to come up with statistics on which artefact is semantically similar to another. Through exploitation of DBpedia information, ratings are assigned to tags again,

---

[1] http://dbpedia.org.

but this time it goes further and tag semantics are compared with each other to try and find the semantic connections among all of them. Ratings are assigned to each established connection, relationships are identified (e.g. parent–child concepts), and for each concept related to another the common semantic links are stored, too. Eventually, a fully fledged semantic network of the domain is formed, while this refinement method manages to construct an invisible hierarchy for each folksonomy, too.

The taxonomy refinement and tag extraction NLP mechanisms have been initially developed as part from the OrganiK project, which proposed a socio-technical approach to KM (Bibikas et al. 2008). More information on how these algorithms and systems have been developed, tested and implemented in KM platforms can be found in the deliverables of the OrganiK EU-funded project,[2] as well as related research detailing the specifics of these algorithms (Christidis et al. 2012a, b). The algorithms focus on using probabilistic topic models (Latent Dirichlet Allocation–LDA (Blei et al. 2003)), in order to attempt and uncover latent topics that can be used to find similarities. The process involves a total of four steps: (a) analyse the items semantically, (b) user behaviour and interaction with content pieces, (c) combination of (a) and (b) towards developing a profile for each item and (d) the recommendation of an item based on the most semantically relevant profile.

The LDA in step (a) is used to derive how topics are distributed over items, and also how words are distributed over each topic, while in (b) the system keeps track of what the user has visited, edited and so on, keeping a related metric. The consolidation of that information being performed in (c) takes the metrics for each topic and user, which is highly dependent on the impact of each metric. Step (d) provides the recommendations and the algorithms account for changes to all these combinatorial profiles that may occur with the introduction of new items, topics, activities, etc. For the currently proposed approach, these systems have been modified to be able and support the multiple folksonomies architecture, as well as to connect to DBpedia as an external semantic knowledge source for step (a). Another addition is that these functions have been modified to also look first for a local copy of DBpedia category label indexes, in case the organisations are hesitant—or in some cases even prohibited—to use online tools and external communication of sensitive or classified organisational knowledge.

The semantic enrichment is the key feature of the proposed KMP as it can efficiently homogenise and give structure to heterogeneous and seemingly unstructured pieces of information recorded in the knowledgebase. The

semantic networks formed underneath unstructured data drawn from posts on social media (e.g. Twitter, Facebook) and online phpBB forums, or structured information from company reports and side wikis to company KM systems, is capable of efficiently and effectively interlinking these knowledge pieces. Towards Web 3.0 and semantic web, this is a huge opportunity for organisations to meticulously gather information from social media and other online sources of knowledge and, through content validation and annotation, incorporate this immense wealth of information as part of organisational knowledge. Proper categorisation with the multiple folksonomies and their subsequent enrichment ensure that all that information is relevant to end-users and that they may exploit it as per operation requirements with the proposed adaptive knowledge retrieval process. The caveat in this case is that end-users may be presented with a seemingly unnavigable information overload, which is tackled mostly through the recommender systems to be discussed in the following section.

5 Recommender systems

The final link in the chain of techniques and mechanisms towards adaptive knowledge retrieval comes in the form of the recommender systems. These recommender systems complement the basic means of retrieving knowledge in the proposed architecture: perform full-text search, browse content categorically via each different blog or Wiki, browse content by a single tag it was annotated with, browse content from latest activity lists and browse content by author. In the approach that we follow here, the recommender engine of the KMP is tightly coupled with the semantics engine; ratings and links of the semantic network behind the tags and folksonomies are exploited to provide accurate recommendations in various forms. It is imperative to stress that while all of the recommender systems described below are enabled by default, an organisation deploying the KMP may disable any number of them to achieve optimal results.

What the recommender systems aspire to achieve is that the information is pushed towards the users and that they do not have to try and navigate on their own through all the potentially intimidating number of resources. Various mechanisms have been employed to ensure that the end-users will receive only the results relevant to their needs, whether through search or browsing. A user looking into a resource that came from company reports may find recommendations of the highly relevant articles from the side wiki or interesting forum/social media posts categorised by the various important aspects of their domain. While the system attempts to push information to users, it does not

---

aspire to make them only passive receptacles of information; rather, through the collaborative filtering processes detailed earlier the users play an integral role in how recommendations will be provided to them, essentially based on how they perceive the information for the domain themselves and how they expect to retrieve it.

5.1 Browsing history

This is a typical recommendation system with a twofold purpose: provide for each user a list of resources similar to those they tend to access, and for each resource currently being accessed to provide a list of resources that other users who accessed this one, have accessed, too. This enables users to more easily find resources that are relevant to their most common interactions with the platform, and therefore to access knowledge that fits into the same categories. This can also be used, however, so that users discover new resources that their colleagues have potentially found relevant to the current resource in the past. This can, in a manner, provide the means of sharing knowledge among colleagues and raising overall expertise among all employees in the organisation.

5.2 Tag similarity

Another typical recommendation system does not make use of the semantic refinement of the folksonomies. It attempts to identify items that share tags with the one currently being accessed and provides a list with all of them to the user. It does not distinguish between different folksonomies, merely focuses on number of similar tags instead. This has the potential to provide very relevant resources to the users in overall content; however, with a much larger knowledgebase it could become ineffective to a degree due to the many content results being provided. Users may also be confused as to the real relevance of the recommended resource because of the lack of some means of categorisation, such as with the domain-specific folksonomies.

5.3 Content similarity

This system entirely forgoes the tag annotations and focuses solely on similarity of the content itself. NLP methods attempt to find linguistic similarities within two items, and a ranking is assigned to the pair. The highest ranked pairs where the currently accessed item is part of are displayed as recommendations to the user. Much like the tag similarity recommender, in a larger knowledgebase this has the potential to cause information overload and cause confusion to end-users. Additionally, it is expected

that accuracy may also be an issue in this case, due to the fact that several pieces of content tend to have similar structure and language (e.g. incident reports in police incidents following similar flow, financial reports covering the same topics with differing measurements). As such, both tag and content similarity can only be evaluated with a large knowledgebase for the organisation the system is deployed for.

5.4 Tag recommendations

Put to action during the content generation process in the KMP, this system primarily supplies the user with a set of tags for each folksonomy based on the text they have provided to simplify and expedite the process; to a lesser extent, it protects a folksonomy from potential contamination. It is context-aware in that it restricts identified terms for each folksonomy only to those that are semantically relevant to it. Users may either disregard the suggested tags or complement them with their own as they see fit, in which case the system recommends tags to them as they start typing. This may prevent tag duplication (e.g. existing term in lowercase, new term in first word uppercase) in the event that the NLP component recommender system failed to extract a term as a potential tag. Naturally, the recommended tags that will be chosen by the users are subject to validation by the expert users.

5.5 Domain-specific content recommendations

The first of the two major contributions to knowledge retrieval personalisation for organisations, this recommender system attempts to offer the resources that are most semantically relevant to the one currently being accessed. The chosen resources are displayed in separate blocks, one for each chosen folksonomy. This categorisation of the recommended resources has been chosen because it possesses the potential to assist users in deciding what may be most relevant path to take towards guiding their investigation of knowledge to the end. It also avoids information overload by categorising resources, so that users know exactly what type of information the proposed knowledge resource is about.

Semantically relevant resources in this particular case are identified as those that primarily have the most common links with the currently accessed one in the underlying semantic network. In case of similar number of links, resources with the highest ratings assigned to those links are preferred instead. The links, but not the ratings, are designed to be visible to a user in the KMP if they wish to study them and understand why they

received these recommendations. Once a folksonomy has been enriched with sufficient terms and linked content, the semantic refinement process may facilitate stability through structure and the emergence of a consensus. Effectively, this may lead to users discovering information, which they could not have known that it is in fact relevant to them.

5.6 Semantic search

The second crucial aspect of achieving personalised knowledge retrieval in an organisational domain comes in the form of a semantically enriched search function. Without the text analysis indexing mechanism and folksonomy refined, this would serve as a simple exact or best match full-text search. With the semantic enrichment, however, it is possible for the search function to attempt and expand search results with resources that do not match entirely the specified term(s), but have been deemed semantically relevant. The semantically similar terms are identifiable much like an exact match would be (e.g. bold font), in order for the user to understand why they are receiving this result. As another precaution to avoid confusion, these results follow after the exact and best matches.

Another property of the semantic search is that it accompanies the free-text search with recommendations that are, once more, categorised by folksonomy. The differentiation from the domain-specific content recommendations is that, in this case, the recommended items are not semantically relevant resources; rather, they are semantically related concepts (i.e. tags). Clicking to enable on one or many of them filters the search results to display only those that the selected concept is semantically relevant to. Content need not necessarily be tagged with the concept, just be related to it from another tag semantically linked with the chosen one.

The benefits of this semantic search and result filtering approach can be immense. Users need not attempt to come up with the one, perfect query to get the result they need; they can start from something they believe is relevant to it and, through an intuitive and simple process akin to the human investigatory behaviour, finally arrive at their destination. This approach also provides counter-measures to alleviate cases of information overload where the user could be overwhelmed with the amount of information they would have to go through. Finally, an added benefit is that the semantic relation depth of the recommended concepts can be specified to the desired number of higher or lower-level concepts that may be included; extremely fine-grained knowledge retrieval can thus be achieved with all these properties.

6 Applications in the HMEs domain: the HOMER KMP

The final architecture of the HOMER KMP is illustrated in Fig. 2, which is inspired by the conceptual design of the proposed KMP for the adaptive knowledge management process (Fig. 1). It will be revisited now in the context of the HMEs domain and the HOMER project, where the proposed architecture has been implemented and evaluated by end-users. The system has been developed as a powerful web application powered by the Drupal content management system, with custom modules developed to deliver the intended functionality. A web-based application was chosen by end-users as the best means of accessing information anytime, anywhere and for collaboration purposes, as well as accessing the knowledgebase within their intranet for security and auditing reasons. A service-oriented approach has been used with restful services for analysing text using the NLP and tag extraction mechanisms written in Java. Communication of the KMP with the online discovery tools for downloading resources and related metadata, and consequently uploading them to the system for validation, is performed in the same manner.

The knowledgebase facilitates the smart processing of knowledge through indexing (text analysis, tags, taxonomies), which is exploited by several mechanisms for providing meaningful knowledge retrieval and recommendations to its users (semantics, browsing history, tags, content). The main interaction of users with the platform, apart from retrieving knowledge via content and searching (semantic search), is the content generation (blogs, wikis, forms) and the potential for discovering new dangers (HME discovery tools). But primarily, users have limited access to the system which is governed by their role in their organisation (role-based access).

In an organisation, employees are assigned specific roles that come with varied rights and responsibilities. This mechanism ensures that this is also reflected in the platform, too. Consequently, the responsibilities of this mechanism lie in the assignment of roles to users, preventing access to each functionality of the system, along with its respective content, to any roles that do not have the authority to access them. As such, this mechanism can leverage the actualisation of the structure of an organisation, or even the structure of a collaborating network of organisations, too, within the HOMER KMP. Several functionalities have also been developed to enforce content classification (i.e. redacted documents using the EU security markings) and to establish a need-to-know basis that reflects the modus operandi of LEAs and other organisations related to the field of HMEs. Through this role-based access offering fine-grained access control to resources and
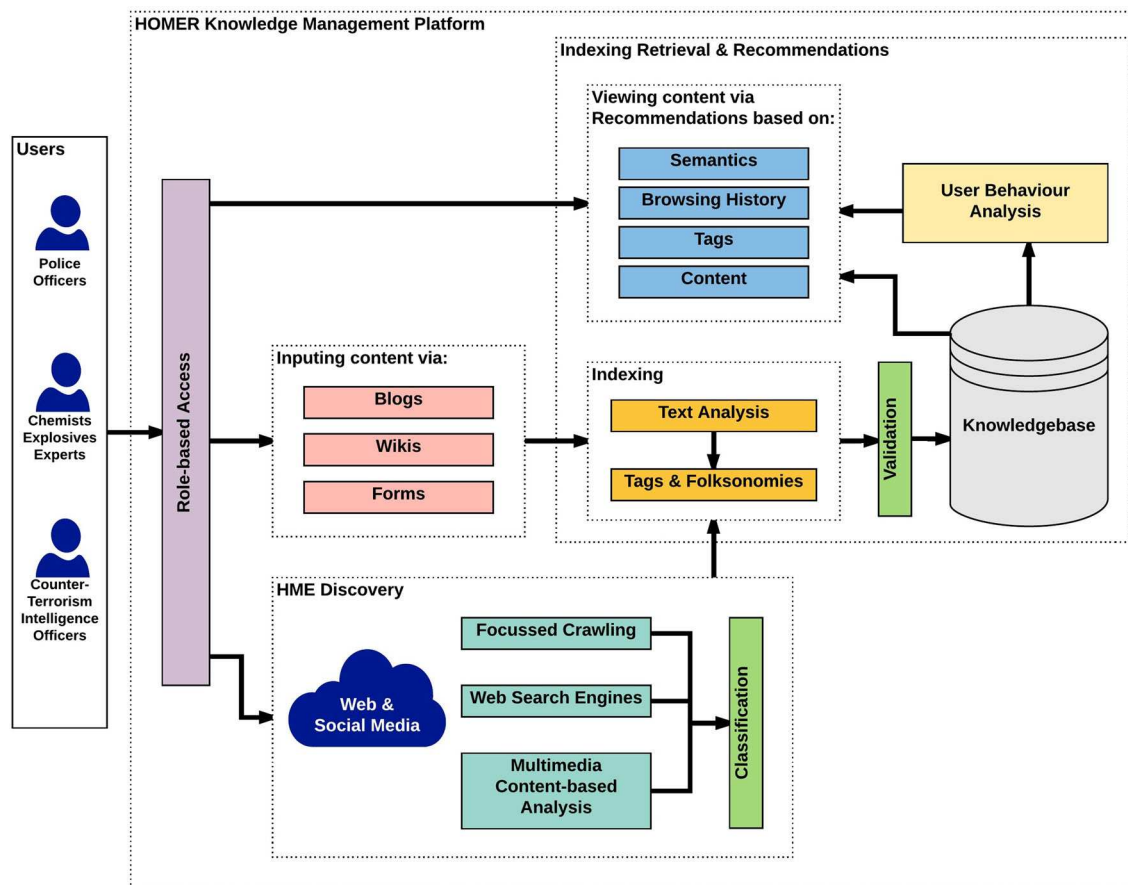
Fig. 2 Architecture of the implemented HOMER KMP

activities on the platform, the users can record, organise, and categorise, retrieve and share knowledge on HMEs.

6.1 Recording organisational knowledge

One of the constituent aspects of a knowledgebase is the recording of knowledge, which this mechanism is responsible for. Recording of knowledge could be a very menial task to be solved easily using tabular or plain-text forms. However, this process is what is employed in traditional database approaches and thus would defeat the purpose of a knowledgebase, where it is essential to plan for the organisation of that knowledge and its meaningful retrieval. Therefore, it is essential for the platform to possess a way to store the content that does not result in the mere recording of data. Two approaches are proposed to this end: the blog (whether plain-text or tabular forms converted to plain text) and wiki formats.

Blogs provide a form of storing information that is more natural to humans, in the context of writing a report versus filling fields in a form, for example. The added value to this option comes with the capability to processing this text using Natural Language Processing techniques (NLP),

which enables the system to avoid the need for constantly updating database schemas as in more traditional methods. Meanwhile, the social aspect exists where other people reviewing that document may effortlessly leave their comments, leading to collaboration towards developing knowledge, as opposed to merely storing information, and it helps promote knowledge dissemination within the organisation, or with collaborating partners through group access control.

Wikis provide the most widespread, popular and efficient form of storing knowledge in a structured and easily accessible manner nowadays. Naturally, this functionality provides an alternative to end-users that prefer certain information to be recorded, and consequently organised and accessed, in this manner. In the case of the HOMER KMP, data pertaining to Explosives, Precursors tied to them, Recipes for making them and the Effects of each such recipe, are organised in a Wiki that has been named HMEs Handbook. Users can thus find all information about known HMEs available to them structured in a hierarchical format easily.

As another form of capturing knowledge related to HMEs, in the current digital era where a wealth of

information can be found online, the HOMER KMP integrates an array of tools that are focused on discovering resources online. These include web search on popular search engines, as well as Dark Web networks such as Tor, a focused crawler that can discover information in online forums and retrieve the posts intact, but it also includes multimedia content-based analysis tools that can determine if video files contain details for preparing HMEs. Information from these tools is fed to the platform by users for review and validation purposes, so that those resources can become part of the knowledge. Social media analysis can also be performed in networks such as Twitter, in an attempt to discover communities and key players involved in HME and terrorism related activities.

Apart from all these heterogeneous types of information being recorded, the HOMER KMP can also accommodate organisational knowledge stored in a structured manner. A prime example of this is the huge number of police reports adhering to a strict format for recording incidents that has been followed for several decades to ensure correctness of the process, and which are being stored in traditional database systems. The KMP is capable of converting structured information such as these reports into unstructured text in natural language, with the consultation of the LEA partners, so that it can properly be processed by NLP algorithms and annotated as needed.

The capacity of incorporating all aforementioned formats of existing and data-mined bits of knowledge in an unstructured format, at the same time managing to turn it into part of organisational knowledge essential for operations, can be highly beneficial towards supporting LEA operations, such as investigatory activities.

6.2 Categorising and organising knowledge

Another aspect of the knowledgebase is the efficient organisation of the knowledge stored in it. Part of this mechanism can also be the by-product of the blog and wiki content generation capabilities. Notwithstanding the contribution of these two features, the most important mechanisms that facilitate this organisation are the analysis of the content and its subsequent annotation with tags. Text analysis is performed using NLP techniques and consulting well-established data models, such as those of DBpedia, but also through tagging that relies solely on HMEs domain-specific folksonomies.

This is also the first step of developing a knowledge retrieval environment specialised for HMEs according to the proposed approach in this work (multiple domain-specific folksonomies). To this end, three folksonomies have been employed that have emerged through discussions with LEA end-users and domain experts as the most important ones: Precursors (chemical components used to make HMEs), Construction (tools and materials used to create Improvised Explosive Devices—IEDs) and Incidents (people, places and organisations involved in HME-related incidents).

Firstly, by portioning the HME domain into the appropriate sub-domains of Precursors, Incidents and Construction, we can narrow down the vocabulary that users may employ to tag resources for each one and avoid confusion. Consequently, this enables multi-faceted knowledge representation for the domain and can be exploited by knowledge retrieval mechanisms (semantics, recommendations) to offer more accurate resources to the investigators in terrorist incidents and potential connections. Nonetheless, the introduction of multiple folksonomies users must utilise during content input and may potentially prevent them from utilising the knowledge recording process to its fullest extent due to being a cumbersome process. This is countered with the utilisation of tag recommendations based on a hybrid recommender system using semantics and NLP techniques. Essentially, users receive the most important terms extracted from the current they are annotating and categorised by each folksonomy, with the extra advantage of minimising problems with tag meanings and synonyms. This is further enhanced through content validation, as discussed earlier.

All these counter-measures required by the second step (protect folksonomies) are thus employed in the HOMER KMP here. Additionally, according to the second step, semantic enrichment of the desired folksonomies describing the domain is also required in the system. As per the design of the system, DBpedia is consulted for each of the aforementioned taxonomies and the proper categories are chosen so that the refinement process provides the best semantic network in the end. To this end, the Precursors folksonomy is assigned the Chemical Components DBpedia category, the Construction folksonomy is assigned the Tools and Materials categories in DBpedia, and the Incidents folksonomy is assigned the People, Places and Organisations categories in DBpedia.

Semantic properties for each term in these folksonomies are derived directly from these categories only, and the algorithms develop the semantic network separate for each single folksonomy beneath by using the relationships among those properties. The information is drawn from tuples retrieved from DBpedia indexes in the form of Spotlight indexes, by which the algorithm can process and determine the proper category a term belongs to, as well as the rating of its relationship to other terms as described earlier.

6.3 Knowledge retrieval for HMEs

Knowledge retrieval is the key to leveraging organisational competence in operations because it is tied to the

identification and accessing of knowledge with regard to HME threat prevention, response and further investigatory activities. Users may retrieve the knowledge they desire through all the popular, intuitive functionalities of the web-based HOMER KMP: browsing entries and tags, reading content and recommended content, or searching for resources online and within the knowledgebase.

Recommendations are the core component of the HOMER KMP that assists the knowledge retrieval tools. The underlying mechanisms belonging to this component can exploit the organisation of the content through tags and folksonomies, use the NLP results from content analysis and record the behaviour of the users to bolster precision. These recommendations appear as a list of suggested posts for each different sub-domain of the HMEs, whenever the user accesses any document-type resources (e.g. Wikis, Blogs). When the user is searching for resources related to their desired terms, however, the recommendations are in the form of a list of concepts the users may wish to explore further and part of the semantic search features. Consequently, the user is provided with additional knowledge to explore what is bound to be very valuable in their investigation in either searching or browsing activities, while at the same time avoiding information overload by categorising everything per the portioning of the domain, hence allowing users to understand what area of the field this information is related to and how.

The resulting recommended resources in both browsing and semantic search can be any type of resource that the HOMER KMP is capable of recording. A user may read, as an example, information on a specific explosives precursor from the HME Handbook side wiki. They receive recommendations in the Construction category in other side wiki information regarding the construction of explosive devices using this precursor, police reports regarding past attacks and organisations using the specific precursor in the Incidents category, and relevant online resources discovered through crawling forums and dark web pages detailing preparation instructions for new recipes in the Precursors category. The user may then take control of the investigation and steer it to the direction they want, such as investigating the resources of incidents to find connections among organisations and precursors, or social media resources to determine which communities seem to be mentioning such precursor elements, even to expand their knowledge by reading resources in the HMEs Handbook wiki.

6.3.1 Recommender systems

In the approach that we follow in the HOMER KMP, the recommender systems are tied closely to the semantics enrichment functions. The ratings and links that are the product of the semantic network formulation and rest beneath the tags and folksonomies are brought together to provide accurate recommendations in various forms. The browsing history is a basic and typical recommendation system with the purpose of giving to each user a list of resources like those they tend to access, and for each resource currently being accessed to have a list of resources that other users who accessed this one, have accessed, too.

Tag similarity is another basic recommendation system; in the HOMER KMP, it does not make use of the semantic refinement of the folksonomies. What it focuses on is to attempt and identify items that share tags with the one currently being accessed and provide a list with all of them to the user. It does not account for the terms belonging to another folksonomy; rather, it only centres around the number of similar tags between two resources. Content similarity recommendations, on the other hand, does not account for the tag annotations and folksonomies, at all, and instead relies on similarity of the content itself. NLP methods attempt to find linguistic similarities within two resources being compared, and a ranking is assigned to each pair. The highest ranked pairs where the currently accessed item is part of are displayed as recommendations to the user.

Tag recommendations are used during the content generation process in the HOMER KMP to protect the folksonomies and expedite tedious or otherwise cumbersome tasks for end-users. It supplies the user with a set of tags for each folksonomy based on the text they have provided to simplify and expedite the process, extracted from the NLP and compared to the semantic networks for each folksonomy separately; context-awareness is the prime benefit in this case. Users may either discard the suggested tags or add to them with their own as they see fit. Figure 3 presents an example of the tag extraction and annotation system at work. The text is being processed by the NLP algorithms, and the user is presented with a collection of terms to annotate the content with. The user can click on tags to add them or again to remove them, or type in their own to complement the ones already suggested to them. General tags tend to present more results as they are not restricted to specific semantic categories related to the domain.

Domain-specific content recommendations are the most important of all the recommender systems employed. The links, but not the ratings, are designed to be visible to a user in the HOMER KMP if the users decide to inquire about why they received these recommendations through mouseover the recommended resource. This can effectively lead users to the discovery of new knowledge within the system that they could not have known it is, in fact, relevant to them. Figure 4 demonstrates a case where the user is browsing a specific recorded incident and they are presented with the domain-specific content recommendations

Fig. 3 Domain-specific and general tag recommendations at content creation (tag extraction)

to the right side of the text. The content suggested for the Precursors category, for example, includes the information on the ammonium nitrate precursor, the ANFO explosive and Terrorist Group A, who have been using such explosives and precursors.

On Construction, it includes content regarding Electronic Detonator, components of which are included in the current content, and again Terrorist Group A has been found or be using such components, so it is also present as a recommendation here, too. That could be one of the important pieces of knowledge pushed towards the user, who can focus on investigating information on this group and navigate through other recommendations there, probably regarding connections to individuals and/or places (the Incidents category). As a final note, the user can mouseover a single recommended content item investigate the semantic links connecting it to the one they are currently accessing (Fig. 5).

6.3.2 Semantic search

The second integral mechanism for allowing expert knowledge retrieval on the HMEs domain is the semantically enriched search function in the HOMER KMP. It follows the principles outlined earlier with regard to the proposed KMP, but now there are the three categories of concepts based on the Incidents, Precursors and Construction folksonomies. The LEA investigators and intelligence officers can navigate through the recommended concepts for each category to narrow down their search results through filtering, or to even discover important and obscure connections.

An example would be to start by searching for a chemical element and then filtering the organisations operating in the users' country and the specific type of IED construction materials, so that they can determine what type of terrorist organisation could be involved in a new incident. Alternatively, sharing the knowledgebase with partners in other countries could help them identify organisation formerly active only in their partners area having now expanded in their own, bolstering collaboration capabilities within counter-terrorism initiatives. Figure 6 demonstrates a semantic, free-text search on a specific precursor of interest to the investigation. Below are the direct results to these searches, with content items mentioning it or annotated with it in order of relevance. To the right again the system presents a few higher- or lower-level semantic concepts related to the search term—this precursor has been used in a Dublin bombing (Incidents category), and Ireland is a higher-level concept to Dublin in
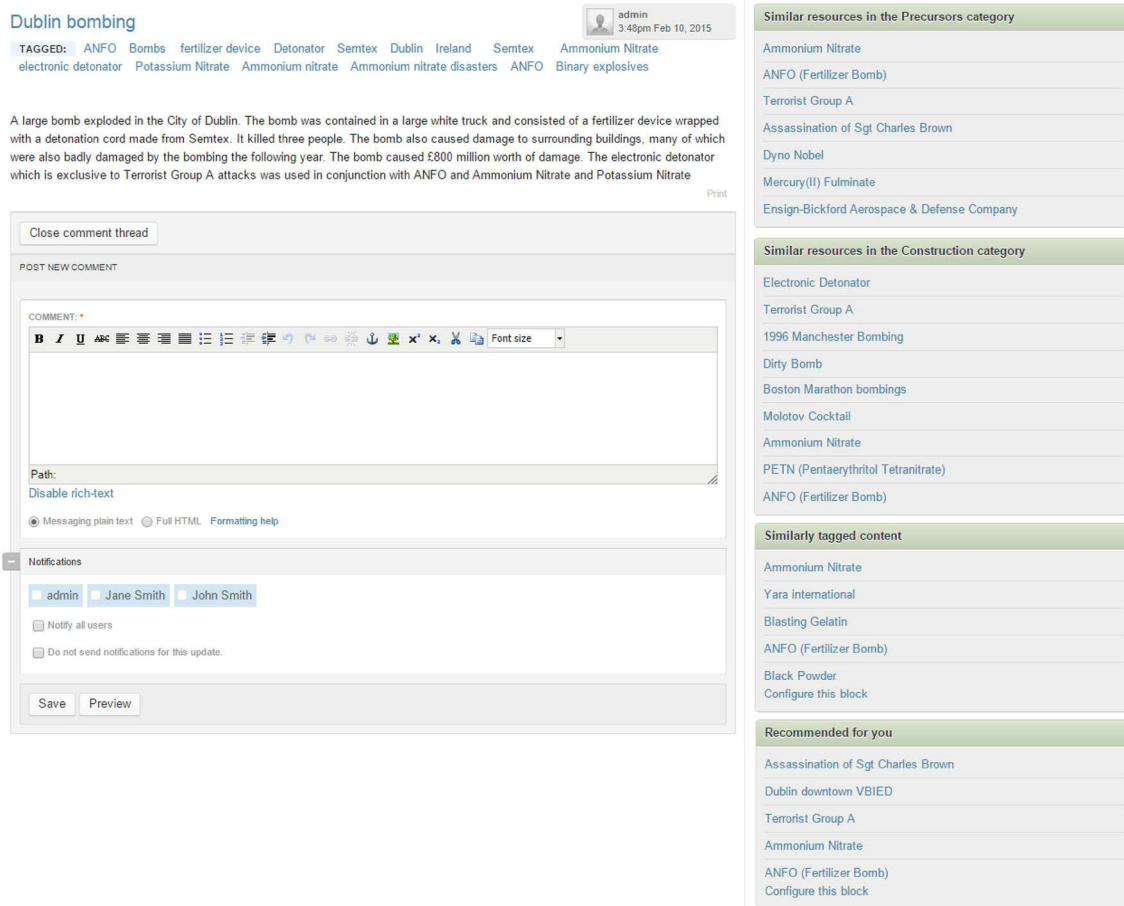
Fig. 4 Examples of related content in domain-specific and general recommendations
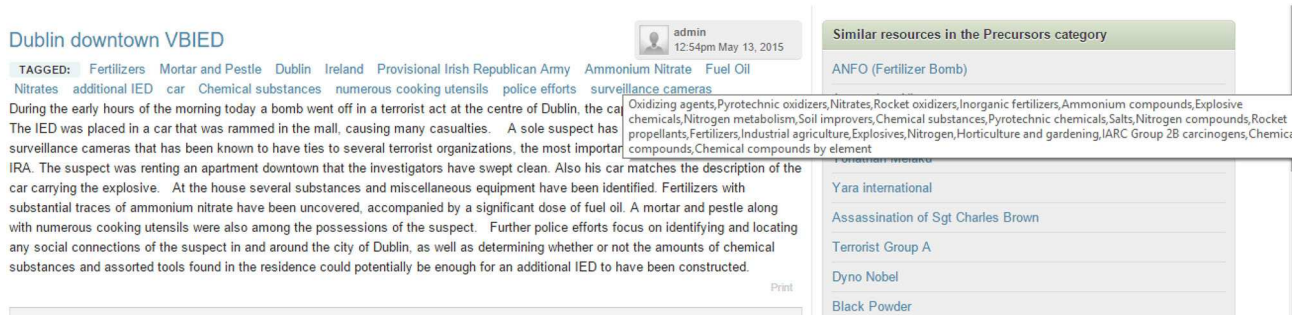

Fig. 5 Semantic links between current content item and recommended resource on mouseover

this case. Users can click on one of these terms, and they will be presented with a list of the resources semantically relevant to this term.

6.4 Sharing organisational knowledge

The HOMER KMP can be deployed in several scenarios, and they vary depending on organisational requirements with regard to sharing information outside the organisation with collaborating parties. Two major choices have been identified that the organisation should make when they decide to utilise the system: (a) if the organisation aspires to utilise the system for cross-agency collaboration and (b) if mobile access should be enabled. This section will briefly explain what are the requirements and additional options in each case. The underlying principle of communication in any case remains the same: a client–server architecture where the KMP is hosted on a single server

Fig. 6 Semantic search sample results in an instance of the HOMER KMP

and users are connected to it through end-point devices depending on mobility arrangements (i.e. desktop based in organisational location or smartphones, tablets and laptops in remote locations, too). Finally, an important note to make is that the following scenarios can be combined, such as a collaboration set-up without mobility options or an in-house solution with mobility options.

6.4.1 Local access scenario

The KMP can only be accessed by devices that are connected directly to the infrastructure owned by an organisation and limited to a specific location that it alone can support (i.e. a LAN). The IP is not utilised in this case, and this solution can provide an ideally extremely secure option for organisations that wish to reap the benefits of a KMP without reliance on outside sources and only for their own gain. However, this set-up can also accommodate collaborative activities when a single KMP is used on a predefined facility that all partners have agreed to administrate and support. There is an immense impediment to collaboration in this case, due to the need of members of different organisations possibly based on different cities to be physically present at the designated location. Consequently, this has been discarded as a viable solution to collaborative activities.

6.4.2 Remote access scenario (in-house solution)

This solution requires a server system to host the KMP and allows any type of devices to connect to it over wireless or wired, remote or location based. Security risks are naturally higher; however, they can be tackled with conformance to proper infrastructure and software solutions. The ideal combination would be VPN access to the KMP, using hardware encryption-capable client devices and hardware encryption modules on the server, as well as strict authentication procedure policies enforced to end-users.

There is a high degree of versatility in the set-ups, a few of which are:

- A server machine at headquarters and desktop clients within them, as well as within regional headquarters and accessible over IP.
- A server machine at headquarters and mobile devices on the field distributed to tactical search teams and the first responders.
- A laptop acting as a server machine, in the field in remote locations with no wireless network coverage, where the laptop acts as a hotspot (e.g. Wi-fi card turned to hotspot, connected to vehicle bearing powerful antenna) and tactical search teams carrying mobile devices supporting their operations.

6.4.3 Remote access scenario (collaboration solutions)

This scenario follows the principles of the one described above; however, access is now given to the single KMP for users outside the organisation. Confidentiality of information and user Authentication are the most serious concerns in this case. Authentication procedures have been described that work not only for access to the KMP itself, but also for information content through the Access Control mechanisms enforced by the Role-Based Access and associated modules facilitating the required need-to-know access. Participating LEAs and other stakeholders should collaborate on the Authentication credentials distribution level (both VPN and KMP account), as well as on the even more important level of assigning the proper Roles, Permissions and Classification clearances to the users that will be connecting to the HOMER KMP.

The partners should decide on one of the following three solutions, balancing organisational requirements with the efficiency of the collaboration:

- Each partner has own KMP, inviting partners with very limited credentials.

  - Minimal gain in knowledge dissemination.
  - Cumbersome for user to keep track of many credentials.
  - Maximum security for own information.

- Centralised KMP hosting all knowledge each partner possesses.

  - Offers maximum gain in knowledge management and retrieval.
  - Inefficient for partners to maintain own KMP.
  - Raises issues pertaining to national security policies with LEAs.

    - Can alleviate with groups, subgroups and classification mechanisms.
    - Immense effort required in achieving joint administration.
    - Issues with location of information hosting.

- Centralised KMP hosting substantial information from each partner pertinent to collaboration efforts:

  - Feasible for partners to employ own KMP.
  - Compromise among security, administration and knowledge dissemination.

    - Consortium needs to determine what bulk of own knowledge is enough to achieve operational needs, but not raise security issues (i.e. aggregation of own information on central KMP raising alert).

7 Evaluation

The HOMER KMP has been developed to be able and support collaborative activities related to HMEs operations LEAs and other interested stakeholders might undertake. The focus is on being able to easily record all types of knowledge, coming from both existing organisational knowledge and online content from social media and forums, use it to find important information towards threat prevention and response, share it with partners and have it available on the field, too. The capacity of the system to do so, however, is directly related to the efficiency of the semantic enrichment and of the knowledge retrieval processes. Consequently, this evaluation of the HOMER KMP is centred on determining how useful the domain-specific content recommendations and semantic search features of the HOMER KMP are towards bolstering investigatory activities for LEAs in counter-terrorism initiatives.

Evaluation was performed on an instance of the HOMER KMP populated with real data provided by the intended end-users themselves, the Police Service of Northern Ireland (PSNI) and Guardia Civil from the Ministry of Interior of Spain (GUCI). The data used towards this end involved standard operating procedures, HME search guides, and declassified, past incidents drawn from their databases and recorded as part of knowledge on the platform through the proper validation and collaborative filtering procedures by the end-user domain experts. Moreover, it included information from online resources discovered by the end-users through the utilisation of the HME discovery tools integrated in the platform. Such resources included sample posts from related forums and pages, from both the surface and the dark web accessible through Tor onion links, as well as wikis and social media. End-users discovered these resources using the search, crawl and community detection tools, validated the information and then merged it with the rest of the knowledge on the platform. Finally, as a result of the collaboration with the rest of the partners within the consortium, an HME Handbook has been created comprising information of approximately 200 precursor elements, explosives and HME recipes.

Initially, the methodology of the evaluation was focused on utilising quantitative methods and looking for the required metrics such as accuracy and recall; however, it became evident over the course of the planning of the process that this could be challenging. Information contained within the knowledgebase, and specifically the aggregation of it all, was considered classified by the LEA partners. Therefore, in this highly secretive and security conscious environment, the partners were reluctant to perform specific, predetermined queries or even share their

own investigation results in a quantitative manner, as this could be perceived as a risk to national security by the LEAs by non-LEA members who had access to it through the HOMER project replicating such queries.

Eventually, the quantitative approach was abandoned in favour of a qualitative approach, where the participants would perform investigations and carry out their own operational scenarios in-house, thus tackling security concerns regarding non-authorised individuals getting involved. The participants were asked to give their overall opinions on the efficiency and effectiveness of the two core aspects of knowledge retrieval in the HOMER KMP and this adaptive knowledge retrieval process: (a) the domain-specific content recommendations and (b) the semantic search. The participants were asked to provide their comments regarding the process of accessing such information, and the relevancy of the results to their current investigation, in an attempt to elicit precision and recall capacity for the recommender systems in a qualitative manner. Additionally, they were asked to provide any additional comments they had regarding the system and its capacity to support operations, as well as any comparisons to the tools they are currently employing.

The domain-specific content recommendations, for the purposes of this evaluation, were configured by default to deliver five maximum recommended resources that were deemed the most relevant by the system, with the option to expand results and see all the rest of the recommended resources. The semantic search was also configured to present a maximum of five of the most related concepts, whether of higher or lower level semantically, again with the option to view all results. The users evaluating the system included five of each organisation (PSNI and GUCI), four intelligence officers and one technical expert in their own systems for operational support, as well as two chemical experts from the Queen's University of Belfast to comment on explosives-related content and performance of the system alone. The system has also been evaluated towards its capabilities of supporting field operations on multiple fronts, such as the first responders and tactical search teams in addition to investigative work at the office during a real-world simulated scenario and live action exercise carried out by the PSNI in Belfast (Fig. 7).

7.1 Semantic search

The recommended concepts in semantic search features have been found to be extremely useful and relevant for investigatory activities, through filtering of results and identification of new links among parts of the knowledge, especially for the folksonomies of Construction and Incidents. Higher- and lower-level concept recommendations were found to be useful for guiding the search, while the
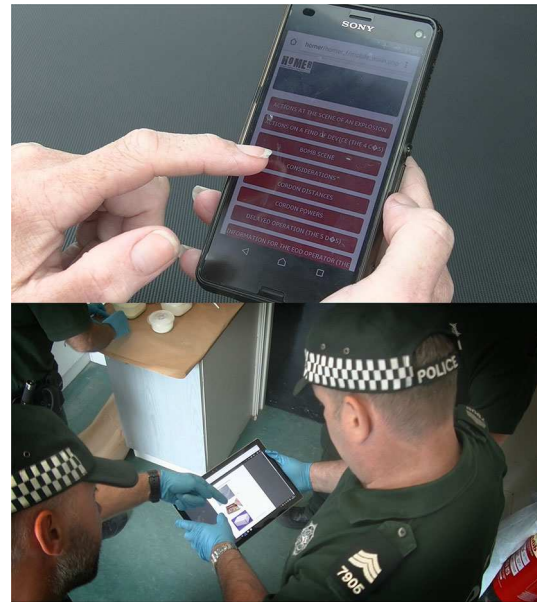

Fig. 7 HOMER KMP mobile and tablet versions field testing at the PSNI real-world exercise

end-users also reported that they are bound to be extremely useful in collaboration initiatives with partners located in other countries that have little knowledge of such connections, as well as their importance in an ongoing investigation. Users also expressed the opinion that having more folksonomies on more sub-domains of the HMEs domain would be highly beneficial to investigatory activities, and that the capability of the HOMER KMP to easily add, remove and refine them would be crucial to the uptake of the system in LEAs.

Concerning the individual folksonomies, the semantic connections among concepts in the Construction folksonomy have been found to properly support investigation into IED components, with the greatest benefit being the linking to past incidents revolving around the search term and the recommended semantically related term in this folksonomy. The added value for LEA operations here was found to be the lack of a need to directly link a new incident to all past ones through direct referencing mechanisms; the semantic network resulting from the enrichment of the folksonomies was more than capable of performing this feat. The same comments also applied to the Incidents folksonomy. However, users also expressed the desire to have this folksonomy, which encompassed information on people, places and organisations, broken down to its constituents. They believed that with the power of having higher- and lower-level semantically related concepts, it could be highly beneficial to narrowing down search results and reaching investigation goals by filtering information across more sub-domains of the HMEs domain.

On the other hand, the precursor folksonomy recommended terms were found to be confusing to the LEA users, but not to the chemist users. This is attributed to the fact that a chemical element (i.e. the DBpedia category used to derive semantic links) can be both a physical or chemical precursor, and some precursors are also explosives on their own. Unlike the case of Incidents where it would be helpful to break down the categories, in this case it was found essential to create totally separate folksonomies to be clear to end-users when a recommended chemical element falls to an Precursor folksonomy or to an Explosive folksonomy. This essentially highlights the problem of properly understanding expected end-user behaviour and requirements outlined earlier, thereby achieving a proper partitioning of the domain the KMP is being developed for to support all operational requirements regarding knowledge retrieval.

7.2 Domain-specific content recommendations

The domain-specific recommendation comments elicited from the users mostly followed in principle those regarding the concepts recommended in the semantic search facilities. Consequently, Construction recommended resources are helpful as they are, Incidents could be further diffused to sub-categories, and Precursors require the distinction between precursor to an HME, and HME or explosive itself. One additional comment in this case, regarding recommended resources and not concepts this time, is that it is essential for the recommendation blocks for each category to have the option to see all linked resources (as the HOMER KMP is now configured) and not merely an acceptable number of resources only (e.g. 5 or 10 maximum). This is because in the domain of HMEs, especially in the precursor/explosives categories, both chemists and LEA users claimed that even the smallest connection to a chemical element can be highly significant to an investigation.

The HOMER KMP has been configured to rank the recommended resources based on semantic links for each individual folksonomy to the currently accessed resource, thereby providing the most semantically relevant resources to users. Evaluation showed a high precision and recall of results for LEA users, who claimed that the topmost five presented were indeed what they could like to look for in their investigations; however, they also claimed that they would go for the option to show all recommended resources and quickly scan them in case something important is hidden in the lower rankings due to small semantic relevance. Chemist users explained that the ordering of results in the Precursors folksonomy is good and that the topmost resources are indeed highly relevant to what is currently being accessed; however, it is still unclear much like with the semantic search functions of whether something is related because it is a precursor or because it has been used an explosive in this case.

In general, regarding the domain-specific content recommendation mechanisms, both classes of users found them to have the precision and recall they would need in their investigatory activities and day-to-day operations. Additional comments to improving the functionality of this knowledge retrieval aspect of the HOMER KMP included the need to add small helpful descriptors on each folksonomy so that end-users completely new to the platform understand what each folksonomy represents and hence minimise training effort and requirements. The number of proposed five topmost relevant results was found useful, but end-users would prefer the number to be ten so that to avoid having to click on expanding results, the reason being that the sensitivity and importance of the HMEs domain is very high and they would like to be more thorough in their investigations. Finally, it was proposed that a better visual representation, or some type of clustering of results, could potentially be very useful to LEA personnel, especially if it could encompass a few statistics.

8 Conclusions

In this paper, we proposed an approach to achieving personalisation for organisations through an elementary, yet effective and efficient, three-step adaptive knowledge retrieval process, which is capable of incorporating assorted heterogeneous pieces of data, even from online sources, into organisational knowledge that can be used to bolster operational capabilities. Key methods employed in this approach are the domain-specific folksonomies, their enrichment with semantics and their protection from common pitfalls of the social tagging approach, as well as the exploitation of the resulting knowledge through expert retrieval mechanisms based on domain-specific recommendations and semantically enhanced search facilities. Adaptivity stems mainly from the capacity to effortlessly add and develop, or remove, folksonomies, thus addressing easily all emerging requirements for organisational operations. Interlinking of unstructured and diverse data, both in content and format, is simplified through collaborative filtering and domain partitioning.

The application of this approach to the intricate domain of HMEs has shown promise with regard to adaptive knowledge retrieval and outlines that the simple, three-step process does indeed require minimal effort to add or remove contextual dimensions as organisational needs change. The approach was also successful in encompassing a multitude of heterogeneous formats for assorted bits of information, turning them into organisational knowledge

capable of supporting operations. Unstructured online resources were effortlessly merged with existing structured organisational knowledge and were accessible as part of meaningful information relevant to case officers.

The approach and its application to the HMEs domain have received positive results according to the latest evaluations from end-users, which include the benefits of an intuitive knowledge retrieval process consistent with human investigatory behaviour, high relevancy of domain-specific recommendations for specific folksonomies, as well as a quick tagging process for promoting the use of collaborative filtering. It has, however, also highlighted some problems that could stem from this proposed approach; specifically, the most important first step of it: errors in deciding on the proper folksonomies that would provide the best results towards knowledge retrieval requirements for end-users may not give enough options to support operations. Nonetheless, the ease of adding and removing folksonomies that describe the domain in the proposed system can significantly alleviate such issues.

Nonetheless, results evaluations with a much more extensive knowledgebase encompassing many pieces of information for all important categories (i.e. incidents, IEDs and organisations) are required to more accurately determine the efficiency of the current parametrisation of the system. A quantitative methodology, in the case that was possible, would provide a much sound and accurate evaluation of the system and its important sub-systems (recommenders and semantics refinement). Future work could revolve around obtaining unclassified content by the end-users to conduct internal tests, complementing those currently being carried out by end-users. Finally, an attempt to bring all three of the steps of this process to the level of a KMP administrative interface is being considered (currently, the second step is missing), therefore providing a framework that does not require code development to achieve organisational goals and tailor the system to emerging needs. This is expected to enhance potential uptake of the system by organisations and to simplify administrative requirements, eventually providing an overall better end-user experience.

References

Bennett PN, Kelly D, White RW, shang Y (2015) Overview of the special issue on contextual search and recommendation. ACM Trans Inf Syst 33(1):1e

Bibikas D et al (2008) Organisational knowledge management systems in the Era of enterprise 2.0: the case of Organi K. in BIS 2008 Workshops Proceedings, 333 pp. 45–53

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Christidis K, Paraskevopoulos F, Panagiotou D, Mentzas G (2012) Combining activity metrics and contribution topics for software recommendations. 3rd International Workshop on Recommendation Systems for Software Engineering, RSSE 2012—Proceedings, pp. 43–46

Christidis K, Mentzas G, Apostolou D (2012b) Using latent topics to enhance search and recommendation in enterprise social software. Expert Syst Appl 39(10):9297–9307

Halpin H, Robu V, Shepherd H (2007) The complex dynamics of collaborative tagging. In: Proceeding 16th International Conference of World Wide Web, pp. 211–220

Iqbal R, Grsywacsewski A, Chang V (2016) Human information seeking behaviour and its impact on personalised information retrieval: an advanced analysis and literature review. Int J Inf Manage

Jabeen F, Khusro S, Majid A, Rauf A (2016) Semantics discovery in social tagging systems: a review. Multimed Tools Appl 75(1):573–605

Loew R, Kuemmel K, Ruprecht J, Bleimann U, Walsh P (2007) Approaches for personalised knowledge retrieval. Internet Res 17(1):49–60

Melville P, Sindhwani V (2010) Recommender systems. Encycl Mach Learn pp. 829–837

Peters I (2009) Folksonomies: indexing and retrieval in Web 2.0, vol. 1. Walter de Gruyter

Rajeswari J, Hariharan S (2016) Personalised search recommender system: state of art, experimental results and investigations.

Robu V, Halpin H, Shepherd H (2009) Emergence of consensus and shared vocabularies in collaborative tagging systems. ACM Trans Web 3(4):1–34

Sharma R, Singh R (2016) Evolution of recommender systems from ancient times to modern era: a survey. Indian J Sci Technol 9(20)

Steichen B, Ashman H, Wade V (2012) A comparative survey of personalised information retrieval and adaptive hypermedia surveys key techniques and impact of personalised IR and adaptive web. i, pp. 1–37

Yao Y, Seng Y, Shong N, Huang X (2007) Knowledge retrieval (kr). In: Web Intelligence, Proceeding IEEE/WIC/ACM International Conference on (pp. 729–735)