Journal of Biomedical Informatics 46 (2013) 160–173

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics

Tien Tuan Dao *, Tuan Nha Hoang, Xuan Hien Ta, Marie Christine Ho Ba Tho

UTC CNRS UMR 7338, Biomécanique et Biongénierie (BMBI), Université de Technologie de Compiègne (UTC), BP 20529, 60205 Compiègne cedex, France

ARTICLE INFO

Article history:
Received 17 March 2012
Accepted 1 November 2012
Available online 10 November 2012

Keywords:
Knowledge-based search engine
Semantic web services
Web-based Human Musculoskeletal System Resources (HMSR)
Ontology
Musculoskeletal system
Biomechanics

ABSTRACT

Human musculoskeletal system resources of the human body are valuable for the learning and medical purposes. Internet-based information from conventional search engines such as Google or Yahoo cannot response to the need of useful, accurate, reliable and good-quality human musculoskeletal resources related to medical processes, pathological knowledge and practical expertise. In this present work, an advanced knowledge-based personalized search engine was developed. Our search engine was based on a client–server multi-layer multi-agent architecture and the principle of semantic web services to acquire dynamically accurate and reliable HMSR information by a semantic processing and visualization approach. A security-enhanced mechanism was applied to protect the medical information. A multi-agent crawler was implemented to develop a content-based database of HMSR information. A new semantic-based PageRank score with related mathematical formulas were also defined and implemented. As the results, semantic web service descriptions were presented in OWL, WSDL and OWL-S formats. Operational scenarios with related web-based interfaces for personal computers and mobile devices were presented and analyzed. Functional comparison between our knowledge-based search engine, a conventional search engine and a semantic search engine showed the originality and the robustness of our knowledge-based personalized search engine. In fact, our knowledge-based personalized search engine allows different users such as orthopedic patient and experts or healthcare system managers or medical students to access remotely into useful, accurate, reliable and good-quality HMSR information for their learning and medical purposes.

1. Introduction

Human musculoskeletal system resources of the human body include the information and knowledge about biological tissues (bone, muscle, ligament, tendon, and cartilage), structures (joints) and their anatomical and functional relationships in normal states as well as in abnormal (pathological) states of the human body. For example, the rectus femoris muscle connects the quadriceps tendon which is attached to the femur bone to move the thigh during working, daily and sportive motions. This information is useful and valuable for the purpose of learning [1]. Another example is a rotational abnormality that could be correlated with the knee and foot positions, femoral anteversion and medical hip rotation through the following functional schema [Gait → Movement → Diarthrosis Joint → Articular Contact → Cartilage → Bone] [2]. This schema means that the gait is influenced by the movement which is acted by the diarthrosis joint. Moreover, its articular contact is characterized by the cartilage which is attached to the bone. Thus, the information about these implicit relationships is valuable for the decision support [2,3]. From the biomechanical point of view, the anatomical and functional relationships between biological tissues and structures can be used to explain the cause–effect relationship of the orthopedic disorders such as clubfeet deformities or rotational abnormalities [2,4,5]. For learning purposes, when a family member is diagnosed with an orthopedic disorder, he (she) and his (her) family members need HMSR (how-to, what-if) about medical processes. Moreover, a clinical expert needs HMSR information about good-quality pathological knowledge and expertise to make appropriate clinical decisions [5]. A healthcare system manager needs HMSR information about medical processes to perform good medical strategy at the macro level [6]. Guest users (medical students, publics) need HMSR information about medical processes, good-quality pathological knowledge and expertise for their learning purposes [7–9]. In fact, there is a great need regarding the useful, accurate, reliable and good-quality human musculoskeletal resources related to medical processes, pathological knowledge and expertise. However, there is no existing system dedicated to such a need.

When having a need of HMSR information, one can use Internet-based resources from mostly used Web search engines such as

* Corresponding author.
E-mail addresses: tien-tuan.dao@utc.fr (T.T. Dao), hoangtuannha@gmail.com (T.N. Hoang), hientpbk@gmail.com (X.H. Ta), hobatho@utc.fr (M.C. Ho Ba Tho).

1532-0464/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2012.11.001

Google, Yahoo, Baidu, Bing, Ask or AOL. However, it is well-known that Internet-based information is unstructured and there is no guarantee about the quality of retrieved results according to a specific request [10]. In particular, for a request of human musculoskeletal system resources, it is important that these resources have to be accurate and reliable. Moreover, the musculoskeletal system resources are relationship-dependant and closely reliant [3]. Consequently, the use of a keywords-based information retrieval system like Google or Yahoo does not satisfy the need of useful, reliable, structured and good-quality information of musculoskeletal system of the human body. In fact, a knowledge-based search engine needs to be developed to take the relationship-dependant character of HMSR into consideration. Furthermore, resources obtained by using a request from a web search engine are volatile, and this is not adapted for learning purposes, which needs a non-volatile storage of retrieved information and knowledge. Consequently, a non-volatile storage strategy has to be applied in such a search engine.

Knowledge-based representation formalisms such as Resource Description Framework (RDF) and RDF Schema, Topic Maps, DARPA (Defense Advanced Research Projects Agency) Agent Markup Language (DAML), Ontology Inference Layer (OIL), and Web Ontology Language (OWL) have become standard frameworks to formalize the information and the knowledge about a domain of interest [2,11–13]. These XML-based knowledge representation languages and standards allow us to define the structure and relationship between information and knowledge through entity and property definitions. In particular, Web Ontology Language (OWL) provides an advanced reasoning level of structured knowledge and information representation. This semantics-rich approach has been used recently in many potential applications such as medical diagnosis system, retrieval information system, or personalized route planning system [3,14–18]. Furthermore, based on these semantics-rich formalisms, web services (i.e. services accessible via the web) and semantics web services (i.e. combination of web service and semantics web technologies) have been developed in a large range of applications from the production system control, the route planning system, the preventive maintenance management in microelectronics field to e-service applications for sale or hotel reservation [19–27]. Consequently, these conceptual and technological formalisms show potential perspectives to develop our knowledge-based personalized search engine in the Biomechanics field.

A significant number of semantic search engines have been also developed [28–33]. New keywords-based [29,34] or map-based [30] or graph-based [35] search strategies have been developed recently to provide user-friendly query approaches as well as to improve the accuracy of the retrieved results. Moreover, the ontology has been used to determine and improve the semantic similarity between retrieved information [36,37]. New algorithms have been also developed for clustering retrieved results [38,39]. Mobile-based communication technologies have been used to facilitate the interaction between the system and end users [40,41]. However, there are no universal frameworks or algorithms for all application domains. The effectiveness and the robustness of these developed systems are domain-dependant. Consequently, specific architecture and algorithms have to be developed in our biomechanics application.

The objective of this present study was to develop a knowledge-based personalized search engine for the web-based human musculoskeletal system resources. Web-based technologies for computers as well as for mobile devices were used to provide a user-friendly and easy interaction between our search engine and the end users.

2. Materials and methods

2.1. Architecture of our knowledge-based personalized search engine

The development of our knowledge-based personalized search engine is based on the semantic web service principle and a client–server multi-layer multi-agent architecture. The search request was considered and modeled as semantic web service request. The developed architecture is illustrated in Fig. 1. Our search engine is composed of three layers. The first layer is the Data Layer which includes our database of web-based content-based HMSR information and a search engine database for user's information and retrieved results. The second layer is the Application Layer which includes our main semantic web service (SWS) agents, an authenticator and semantic web service (SWS) and services managers. The third layer is the Presentation Layer which includes home-machine interfaces for the interaction between our system and the end users. Different types of device such as personal computer or mobile phone, smart phone or Personal Digital Assistant (PDA) can be used to connect to our knowledge-based search engine.

The use case of a service-based search request is presented in Fig. 2. After registering successfully, a user logs into our search en-

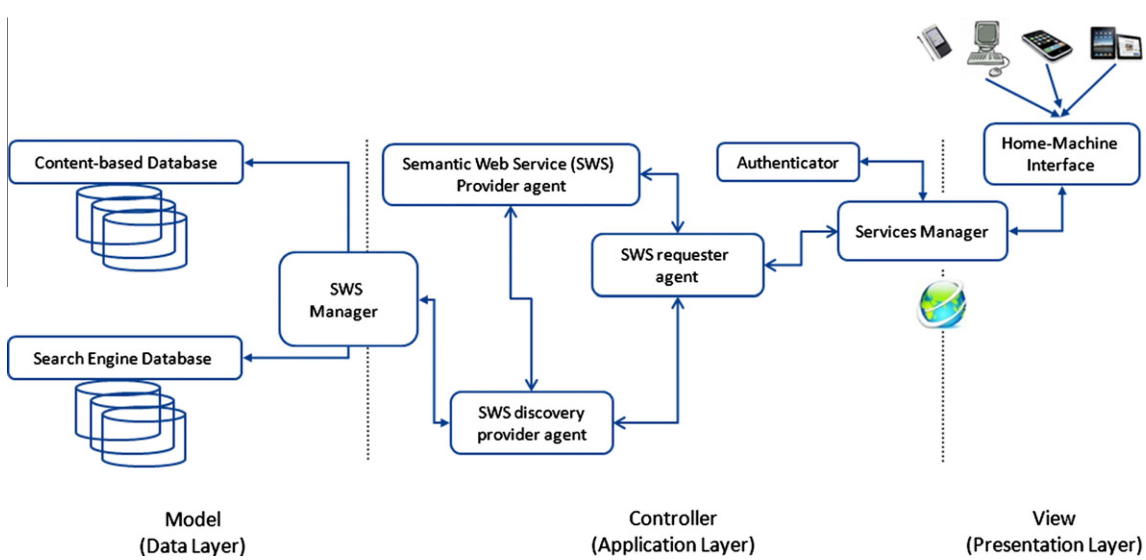

Fig. 1. The client–server multi-layer multi-agent architecture of our knowledge-based personalized search engine.

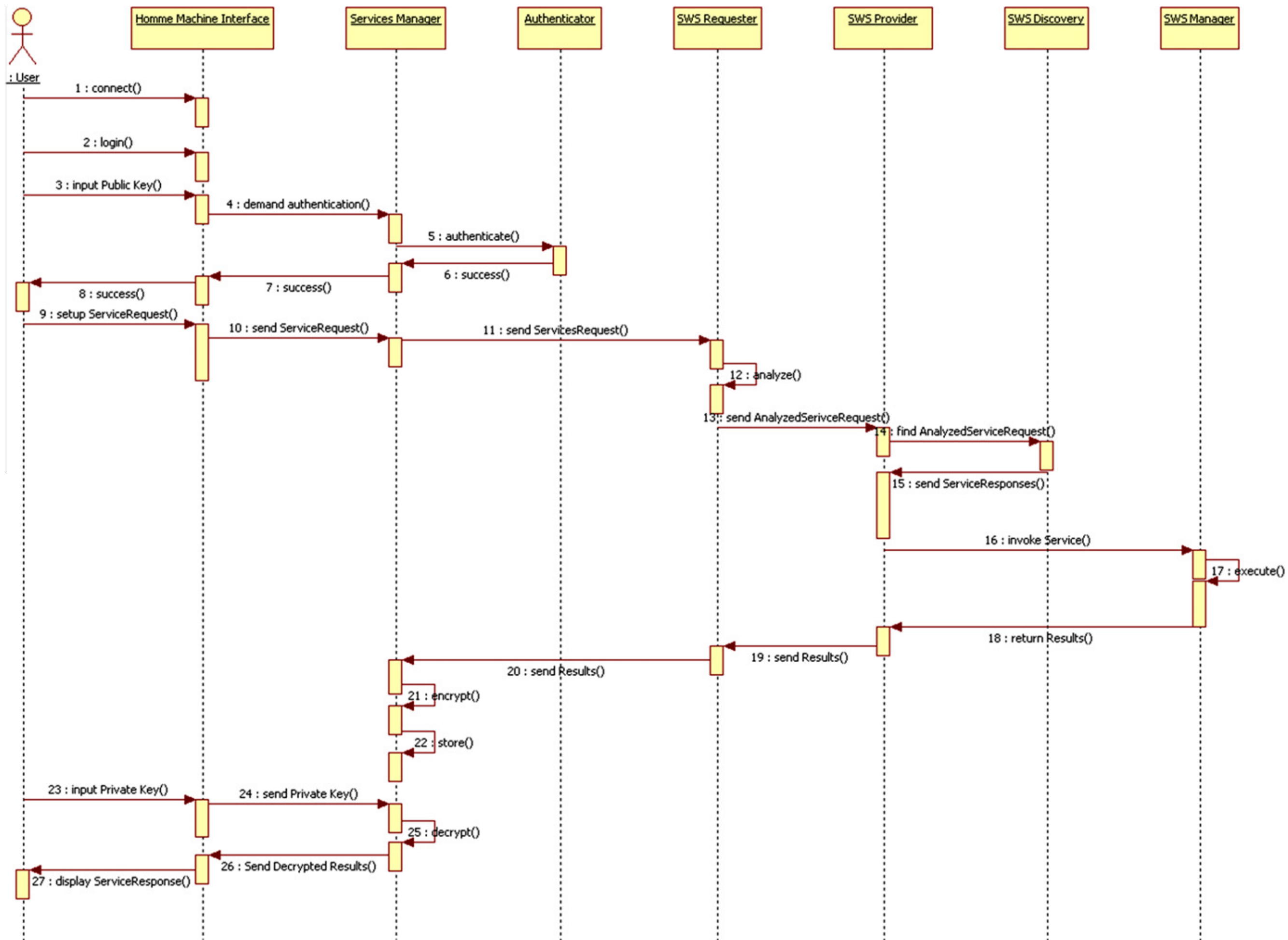

Fig. 2. Use case of a service-based search request.

gine by using his (her) username and password (actions 1 and 2). The authenticator aims to verify the user identification and authentication information (action 5) based on the symmetric key encryption mechanism provided by the hypertext transfer protocol secure (https) protocol (from action 3 to action 8). Then the user parameterize his (her) search request via home-machine-interfaces (action 9) and send it (action 10) to the services manager which transfer it to the SWS requester agent (action 11). The SWS requester agent aims to analyze the request from the end user (action 12) and connect it (action 13) to the involved SWS provider agent. Next the SWS provider agent searches (action 14) the analyzed service request and receipt (action 15) the responses provided from the SWS discovery agent. The SWS provider agent invokes (action 16) the service from the SWS manager which executes (action 17) the service request and return the results (action 18) to the SWS provider agent. The SWS provider agent sends (actions 19 and 20) the results to the services manager which encrypts (action 21) and stores (action 22) the results. When viewing the results, the end user has to provide (actions 23 and 24) his (her) personal private key to decrypt (action 25) the results which are sent (actions 26 and 27) to the user display interface by the services manager.

2.2. Description of our semantics web services and their operational scenarios

The semantic score of our search engine is based on an available ontology named OSMMI dedicated to the musculoskeletal system of the human lower limb [2]. This ontology consists of 14 biological tissues and structures (i.e. *Nervous system*, *Ligament*, *Muscle*, *Tendon*, *Cartilage*, *Bone*, *Limb*, *Posture*, *Support of load, Diarthrosis Joint, Movement, Articular, Contact, Contact of environment and Gait*) which connect to each other by 10 principal anatomical and functional relationships such as *inform, command, attach, compose, act, influence, form, support, create, and characterize*. Each biological tissue or structure part has several sub-parts as the *soleus* is a subpart of the *muscle*. An example of the anatomical and functional relationship is that the muscle is attached to the tendon and it acts on the involved diarthrosis joint to move the bone.

The operational scenario of our search engine ranging from semantic web services translation to the content-based database connection is illustrated in Fig. 3A. Based on our computational ontology developed with the OWL (Ontology Web Language) format, a *Template Ontology* was created using Web Services Description Language – WSDL. Then a transcription procedure using standard logic format OWL-S (a W3C recommendation) was performed to convert this template ontology to the formal description (*Template Service Ontology*) of our semantic web services including search service, e-Update service and e-Newsletter service. The search service aims to request the HMSR information. The e-Update service aims to request the updated HMSR information about some detailed topics. The e-Newsletter service aims to request all new HMSR information.

The request of a user was modeled as a *User Service Ontology* fulfilled by the user's input from the web-based home-machine interfaces (Fig. 3A). Then the SWS requester agent extracts the input

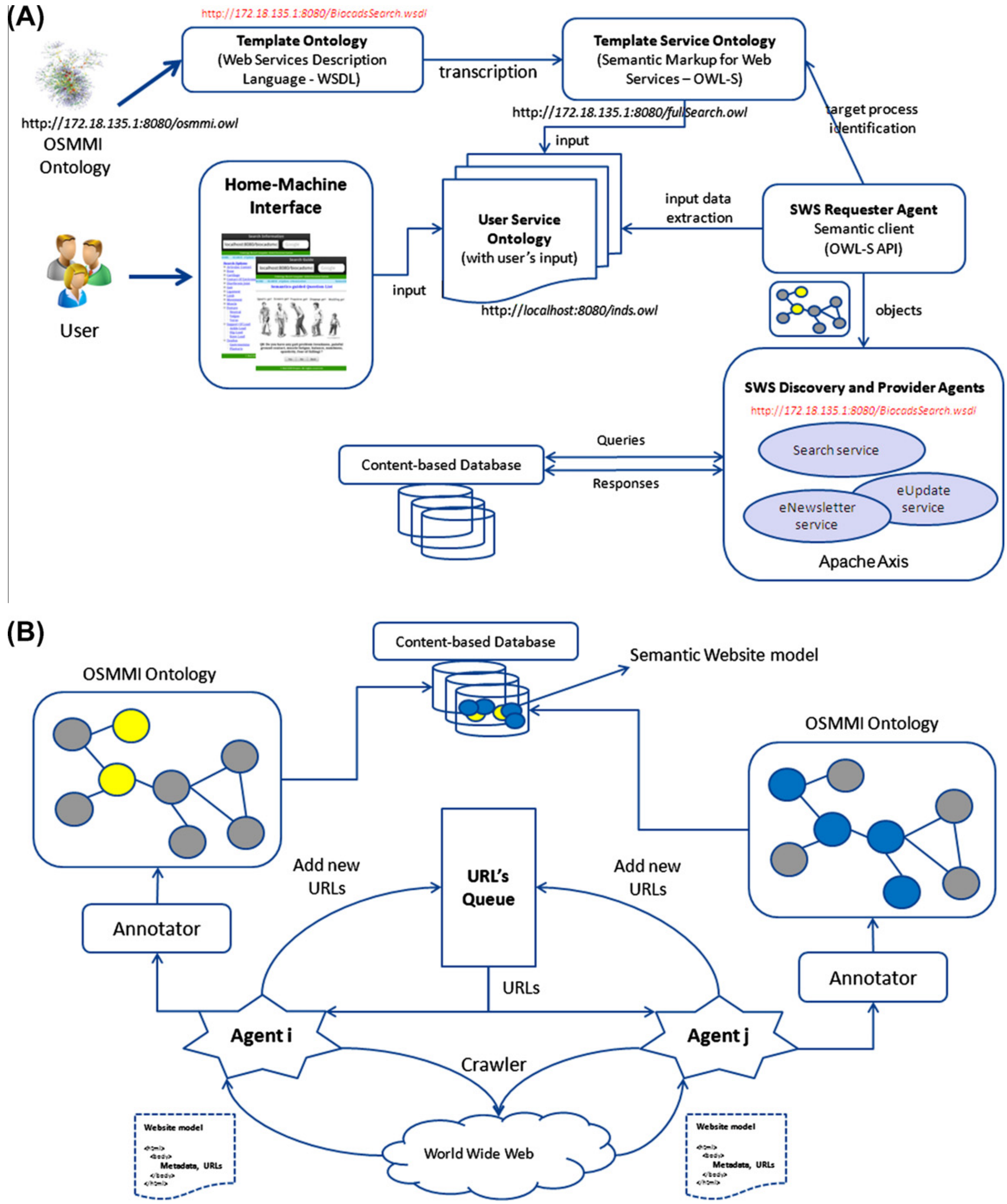

Fig. 3. Operational scenarios of our semantics web services (A) and multi-agent crawler (B).

data from the *User Service Ontology* and identifies the target objects from the *Template Service Ontology*. The SWS requester agent transmits identified objects to the SWS discovery and provider agents which connect to the content-based database for sending queries as well as receiving responses.

2.3. Web-based multi-agent crawler

To develop our content-based HMSR database, a multi-agent crawler was created (Fig. 3B). Based on the initial seed list, each crawler agent collects the HMSR information defined as a website model including Meta data and URLs from the Internet. Then these new URLs are added in the seed list to update this list. Note also that in the case of ontology extension, this seed list will be updated with new URLs related to new entities. In parallel, each crawler agent connects to the annotator to make the annotation of each retrieved website model based on the entities and relationships of our OSMMI ontology. Finally, all of the data are stored in our database as semantic website models.

An example of annotation procedure of a crawled webpage model wP_i is presented in Fig. 4. It is assumed that the website model wP_i contains four entities according to the concepts defined in a simplified ontology E_4, E_5, E_6, E_7 and 3 properties such as P_3 and P_5 of the concept E_5 and P_7 of the concept E_7. Consequently, the semantic web swP_i is stored in the database with all activated concepts and properties (in light blue color).

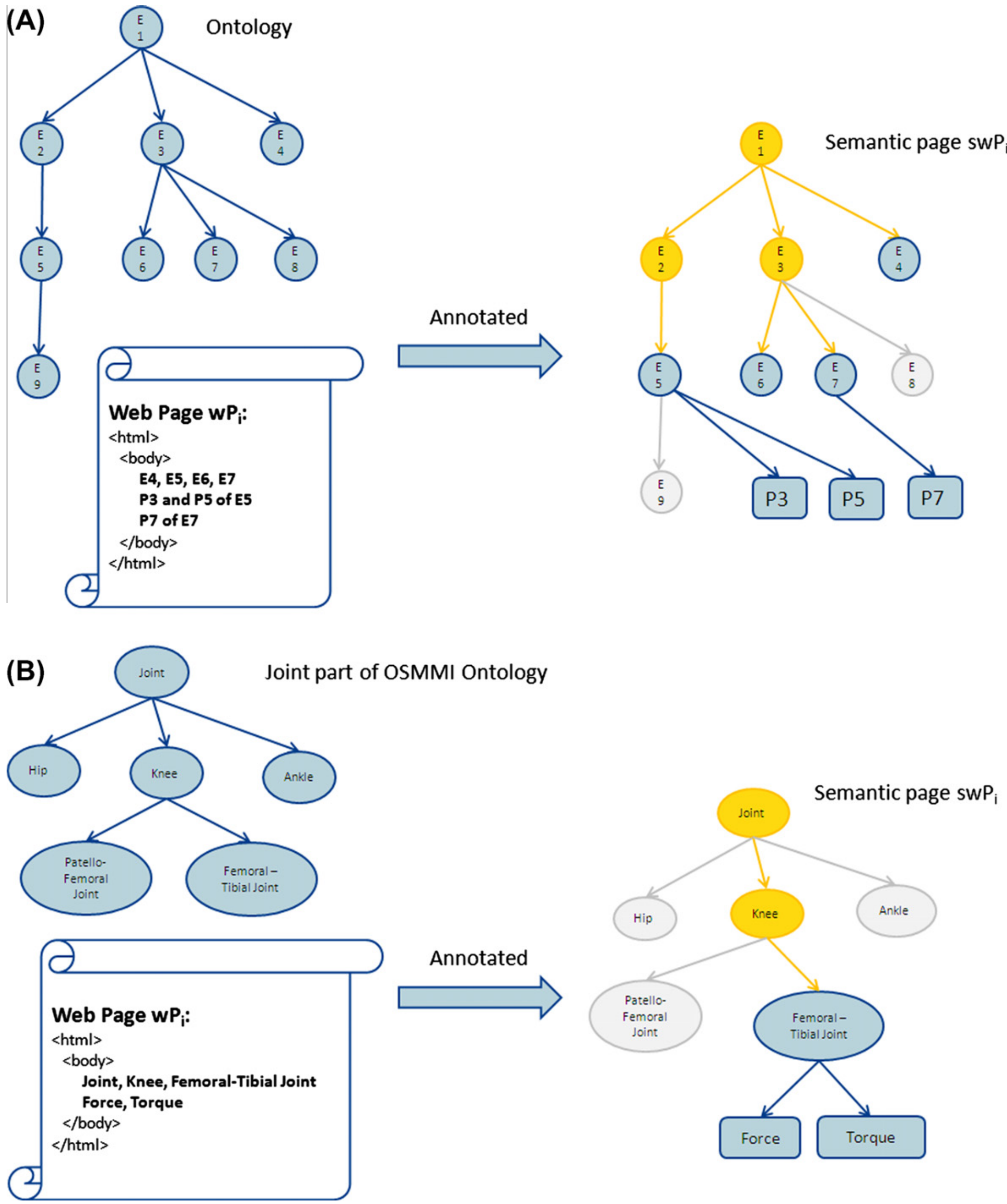

Fig. 4. Example of annotation procedure (conceptual (A) and concrete (B) examples) for a crawled website model wP_i: activated concepts/properties are in light blue color; concepts in orange color reflect their activation due to the direct relationship with an activated concept; gray color reflects the non-activated concepts/properties which are not used for the annotation purpose.

To create the initial seed list, 5-best-first significant website models for each entity defined in our OSMMI ontology were acquired by a biomedical expert from Google Search Engine and Yahoo Search Engine. The best criterion is based on the evaluation of the biomedical expert. The first criterion is based on the ranking order of the used search engine. These website models have to deal with anatomical information-based or knowledge-based websites. Moreover, these website models must have more out-links. Website models such as blog, forum websites, dictionaries, and translator websites were neglected. In the case of ontology extension, new 5-best-first URLs related to each new entity need to be populated by the specific biomechanical expert in the related field of ontology extension to update the seed list.

2.4. Google-based and semantic-based PageRank algorithms

To show the most significant results according to a user's specific request, the set of web-based results from the semantic search request are ordered using Google PageRank algorithm [42,43] if only one concept was selected. If two or more concepts were selected, the retrieved results are ordered using a semantic-based PageRank algorithm.

The Google PageRank algorithm is based on the link structured-based principle [42–44]. A Google matrix based on the web graph structure is used to compute the well-known PageRank score for each retrieved website model. In our system, a standard version of Google PageRank algorithm was implemented.

To develop our semantic-based PageRank algorithm, a semantics-based weighted PageRank score (SPS) was defined. This score is based on two semantic rules. The first shows that the property is more informative (semantic) than the concept and the relationship. The second rule is that the more semantics-based PageRank score (SPS) is greater, the more web page is informative (semantic). This score is computed using the following mathematical formula:

$$\mathrm{SPS}(wL_k) = \sum w_i(L) + \sum 2 \times w'_i(L) \quad (1)$$

where wL_k is a web link, $listC = \{C_1, \ldots, C_n\}$ is the set of concepts/sub-concepts selected by the user, $listP = \{P_1, \ldots, P_w\}$ is the set of properties selected by the user. Each concept/sub-concept C_i or property P_i is defined as a node of our OSMMI ontological graph. $listwP = \{wP_1, \ldots, wP_o\}$ is the set of web links, w_i is the weighted coefficient of link i (named L and denoted as a edge of our OSMMI ontological graph) between two concepts, w'_i is the weighted coefficient of link i between a concept and a property. The weight coefficients w_i and w'_i are computed by the following mathematical formulas:

$$w_i = w_A \times w_B \quad (2)$$

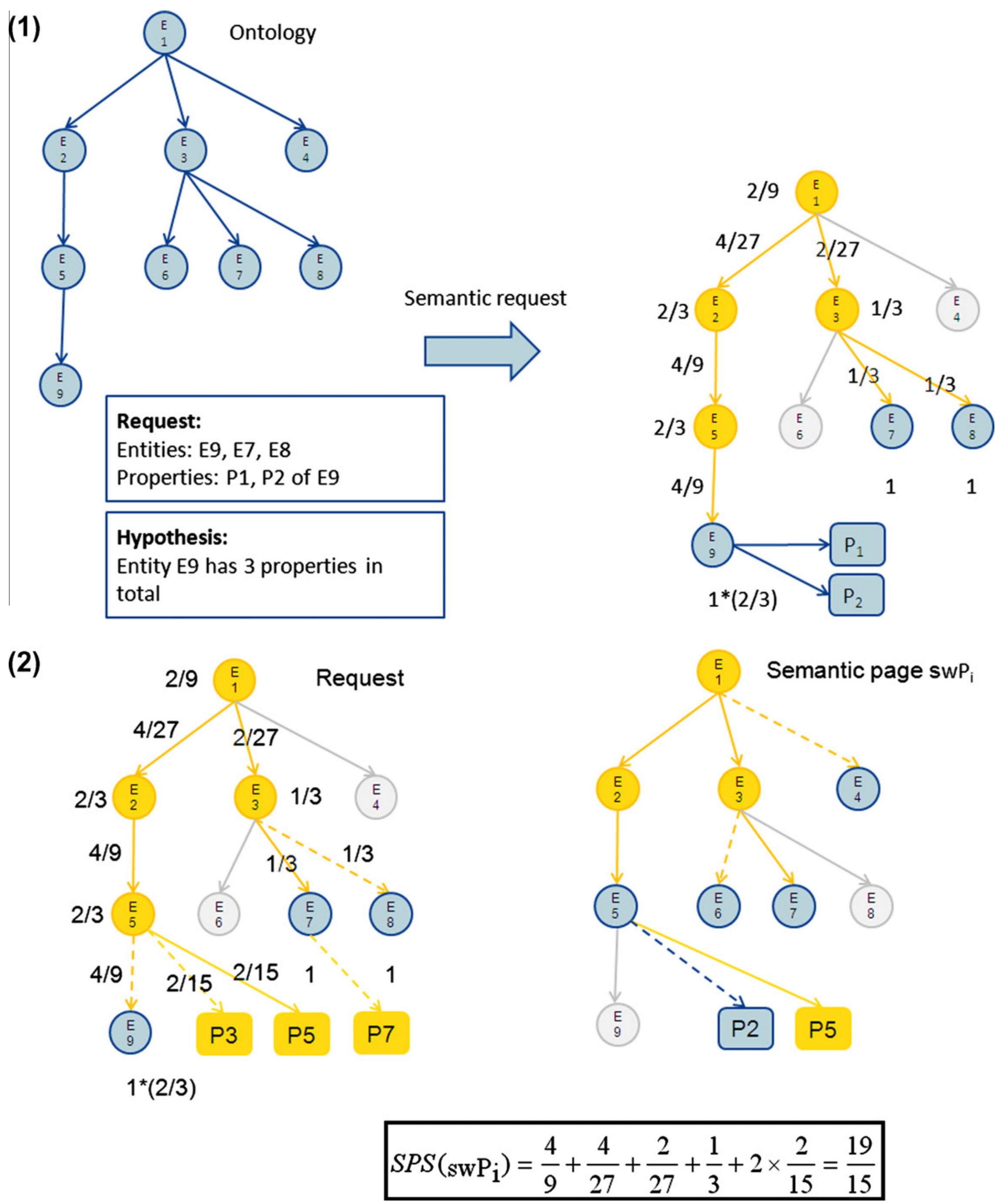

Fig. 5. Two-step (step 1 and step 2) process to compute the semantics-based weighted PageRank score of each semantic page wP_i: non-activated concepts/properties (in gray color and dotted lines) are not used to compute the weight and the semantic PageRank score.

$$w'_i = \frac{w_A}{n} \quad (3)$$

$$w_A = \frac{\max\{W_{B_i}\}}{n} \quad i = \overline{1, n} \quad (4)$$

where w_A is the weight of the concept A, w_B is the weight of the concept B connected to the entity A, n is the number of the properties of the concept A, B_i (with $i = \overline{1, n}$) represents each concept B connected directly to the concept A.

The computing of semantics-based weighted PageRank score is based on two-step process illustrated in Fig. 5. Based on the request of the user with selected concepts ($\in$ *list C*) and properties ($\in$ *list P*), a request-based computing step is performed by using the formulas (2)–(4) (Fig. 5 (1)). Then these results are used to compute the semantics-based weighted PageRank score of each semantic page swP_i stored in the content-based database (Fig. 5 (2)).

Given the Ω is the number of crawled websites models from the Internet; w is the number of annotated website models in our content-based database; n and m are the numbers of OSMMI's ontological entities and properties respectively. Hence the computational complexity of the annotation process is in $O(\Omega nm)$. Regarding the computing of semantic-based PageRank score, the worst-case computational complexity is in $O(nm + wnm) = O((1 + w)nm)$. All retrieved ranked results were ordered using quicksort algorithm which has a computational complexity of $O(w \log w)$ in the average case. Thus, our semantic-based PageRank algorithm can be solved in polynomial time.

2.5. Testing and evaluation

Testing cases with increasing number of properties ranging from 1 to 8 were analyzed. A use case with a specific request regarding an orthopedic musculoskeletal knee pain problem was established. Queries were set up with most meaningfully contents related to the knee pain problem such as knee anatomy (Q#2, Q#3, Q#4), knee contact (Q#1) and knee kinematic function (Q#5) through the length of hyaline cartilage, the length of quadriceps muscle, the length of Achilles tendon, the knee reaction force, and the sagittal rotation of the knee respectively. The user applied our knowledge-based activation principle to refine the query. The precision metric based on the 100-first relevant and irrelevant retrieved websites was computed to evaluate the performance of our semantic search engine. Each retrieved result was reviewed and assessed by a biomechanics expert having around 20 years of qualifying work experience on the biomechanics domain, especially on the musculoskeletal disorders to determinate its relevant or irrelevant character.

2.6. Technologies

Java-based technologies were used to develop our knowledge-based personalized search engine. All algorithms were implemented in Java. Home-machine interfaces were developed using JSP and Aptana Studio 3 (JavaScript and HTML). MySQL was used as database management system. OWL-S 1.0 API (Ontology web language Service), Apache Axis 1.4, WSDL (Web Services Description Language) were used to develop semantic web services. Crawler4J was used to develop our multi-agent crawler.

3. Results

3.1. Semantic web service translation and mapping

A translation and mapping case of our semantic web services from our OSMMI ontology is illustrated in Fig. 6. The concept "muscle" with the instance "adductor magnus" is presented in OWL, WSDL and OWL-S formats respectively. This concept includes different morphological, mechanical and physiological properties such as muscle stress or muscle length or muscle force. defined as *rdf data* in OWL format. By using WSDL format, each property was mapped into a wsdl:part which was mapped into a grounding:owlsParameter in OWL-S format. All remaining concepts were mapped from the same principle.

3.2. Identification of selected entities and properties for semantic web service

Based on the operational scenario of our semantic web services (Fig. 3A), our SWS requester agent must identify the semantic objects selected by the user and send them to the SWS discovery and provider agents for the service analysis and invoke purposes. An identification procedure of retrieved objects performed by the SWS requester agent is presented in Fig. 7. It is assumed that the user selects the adductor magnus muscle and its beta orientation angle and force as properties (values in red color[1]). Based on our mapping process and implemented identification functions, this selected object with its properties were localized and extracted.

3.3. Multi-agent crawler and content-based database

The interfaces of our multi-agent crawler are illustrated in Fig. 8. The number of parallel agents can be parameterized according to the network traffic of the testing platform. Seed list was created and new URLs are annotated and added in our content-based database. Our multi-agent crawler can be stopped and restarted when necessary without loss of information.

3.4. Home-machine interfaces

The main interfaces of our web-based search engine for personal computer were developed. The end user can perform semantic web services such as search request, e-Update and e-Newsletter services. The end user can also view his (her) historical service requests or connect to the mobile-based interfaces of our search engine. Note also that unit and integration tests were performed to verify the correctness and robustness of all implemented functions. An example of the interface of semantic service-based query is illustrated in Fig. 9. The *User Service Ontology* is represented in the form of tree representation. The user can walk the tree by the computer mouse to select the concept and its desired properties. Then the user can select the default values as well as enter the new values of selected properties to refine the search request. Contextual help was provided for explanation purposes.

In addition to the standard service request, a dedicated semantics-guided service request was developed. Web-based interfaces for the mobile device of this semantics-guided service request are illustrated in Fig. 10. The end users fulfill a yes–no questionnaire. The selected responses are used to affine and reduce the number of target concepts (Fig. 11). This semantic-based question-driven search process is based on the semantic relationships defined in our OSMMI ontology and our biomechanics knowledge [3]. The color code was used to illustrate the activated target concepts according to the "yes" response of a question. For example, the Posture and Load concepts are activated if the question 5 concerning the support has the "yes" response.

3.5. Testing cases and evaluation

The retrieved results of two test cases are illustrated in Figs. 12 and 13. The first one related to the select of a single concept

[1] For interpretation of color in Figs. 4 and 7, the reader is referred to the web version of this article.

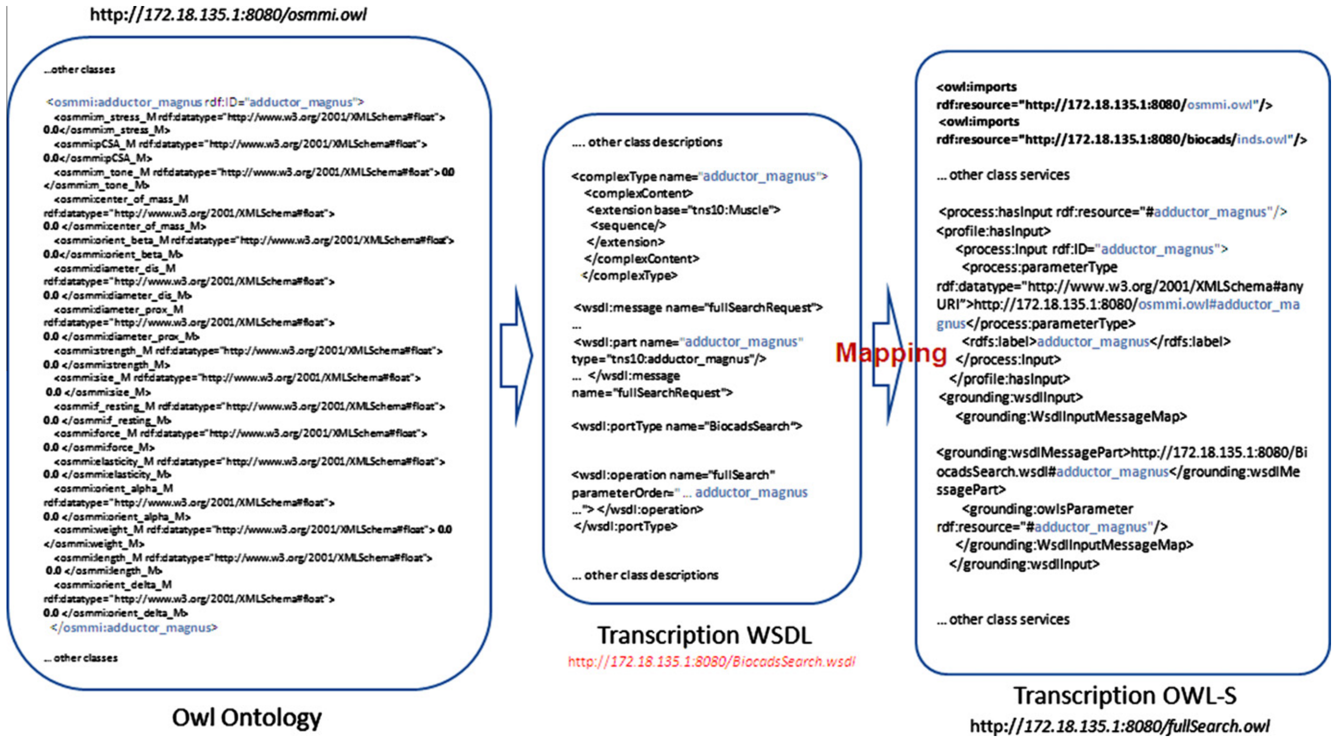

Fig. 6. Translation and mapping of our semantic web services: case of the adductor magnus muscle and its orientation angle and muscle force.

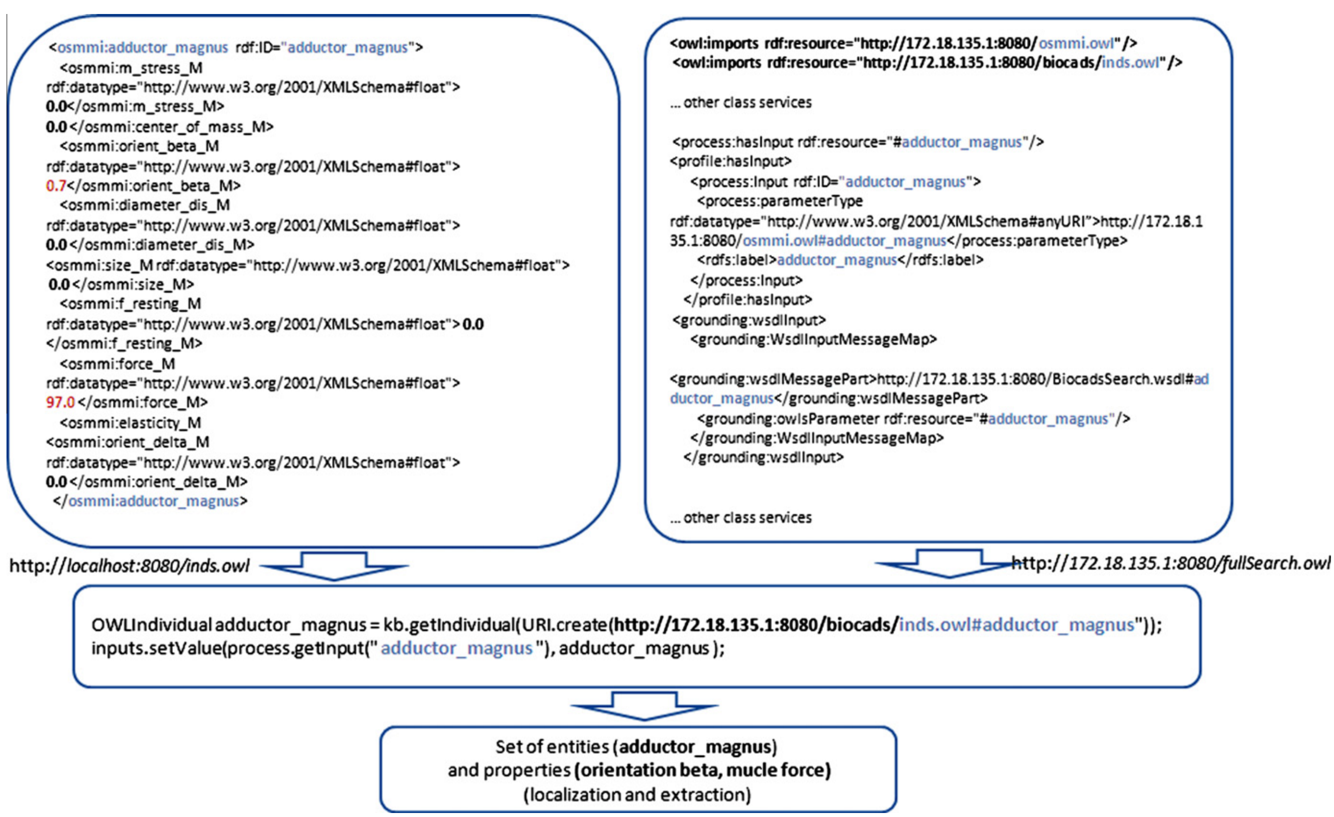

Fig. 7. Localization and extraction of identified objects: case of the adductor magnus muscle and its orientation angle and muscle force.

"Bone", so the retrieved results are ordered by using Google-based PageRank score. The second one deals with 2 selected concepts such as "Bone" (with elasticity as selected property) and "Cartilage" (with length as selected property), so the retrieved results are orders by using our semantics-based PageRank score.

To show how our semantic search system is able to provide accurate, reliable and good-quality HMSR information for a specific request, a search use case scenario was defined as follows: one 7-year-old girl has a pain on her knee, after the doctor visit, her father was informed that her knee problem can be occurred due

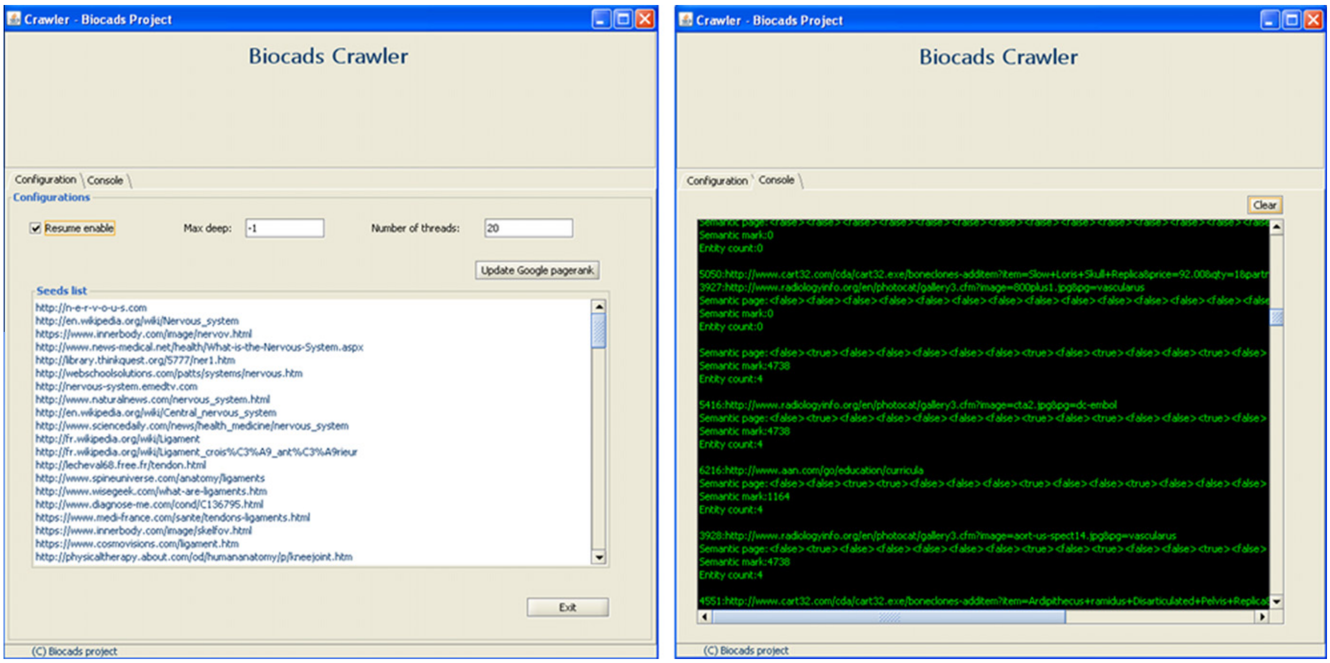

Fig. 8. Interfaces of our multi-agent crawler: seed list (left) and new retrieved website models.

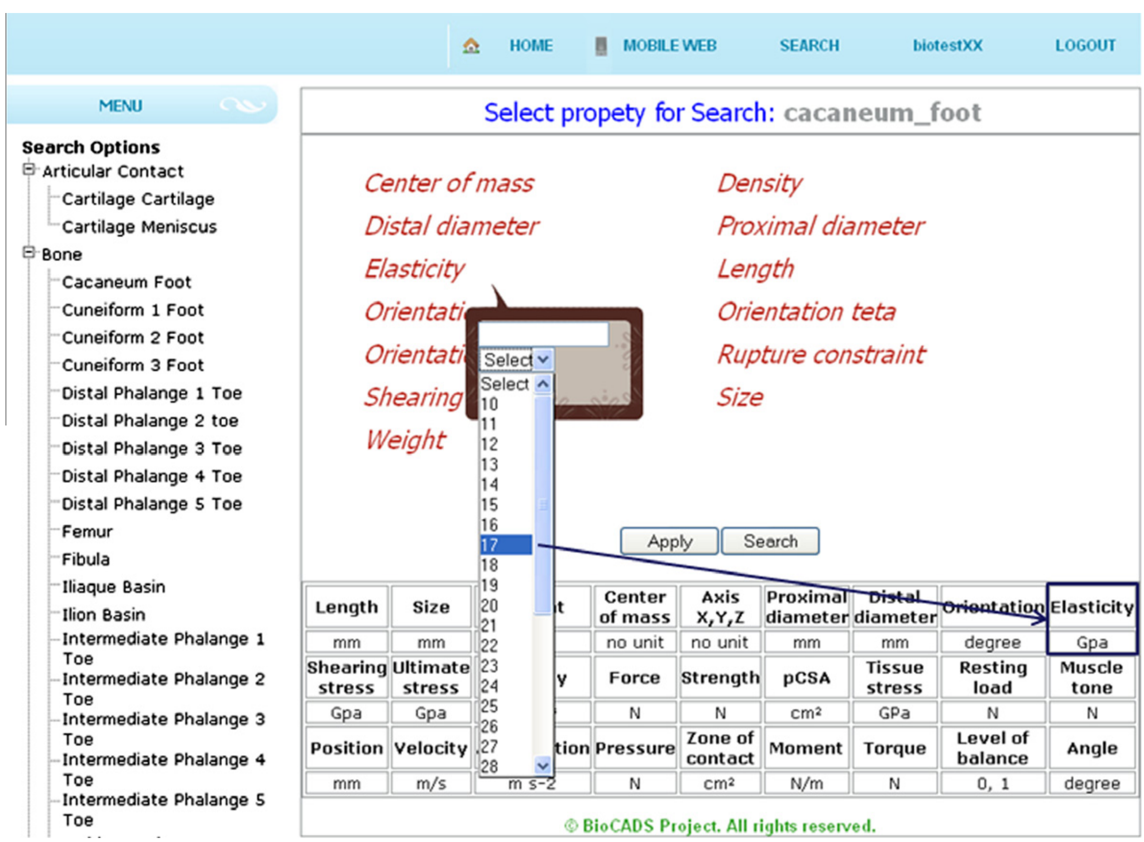

Fig. 9. Query set up for a semantic web service request.

to the ligament rupture. Her father worries about the medical process of his girl and he needs related HMSR information. If he uses a conventional search engine with the keywords "knee pain problem", he is submerged in a huge quantity of information without guarantee of quality and accuracy of these unstructured information. Inversely, when using our semantic search system, he is guided to refine his query by answering some questions based on our activation principle described in Fig. 11 leading to the disambiguation of his query set up (Table 2). Moreover, retrieved information are provided in a context-based manner (e.g. "problem"

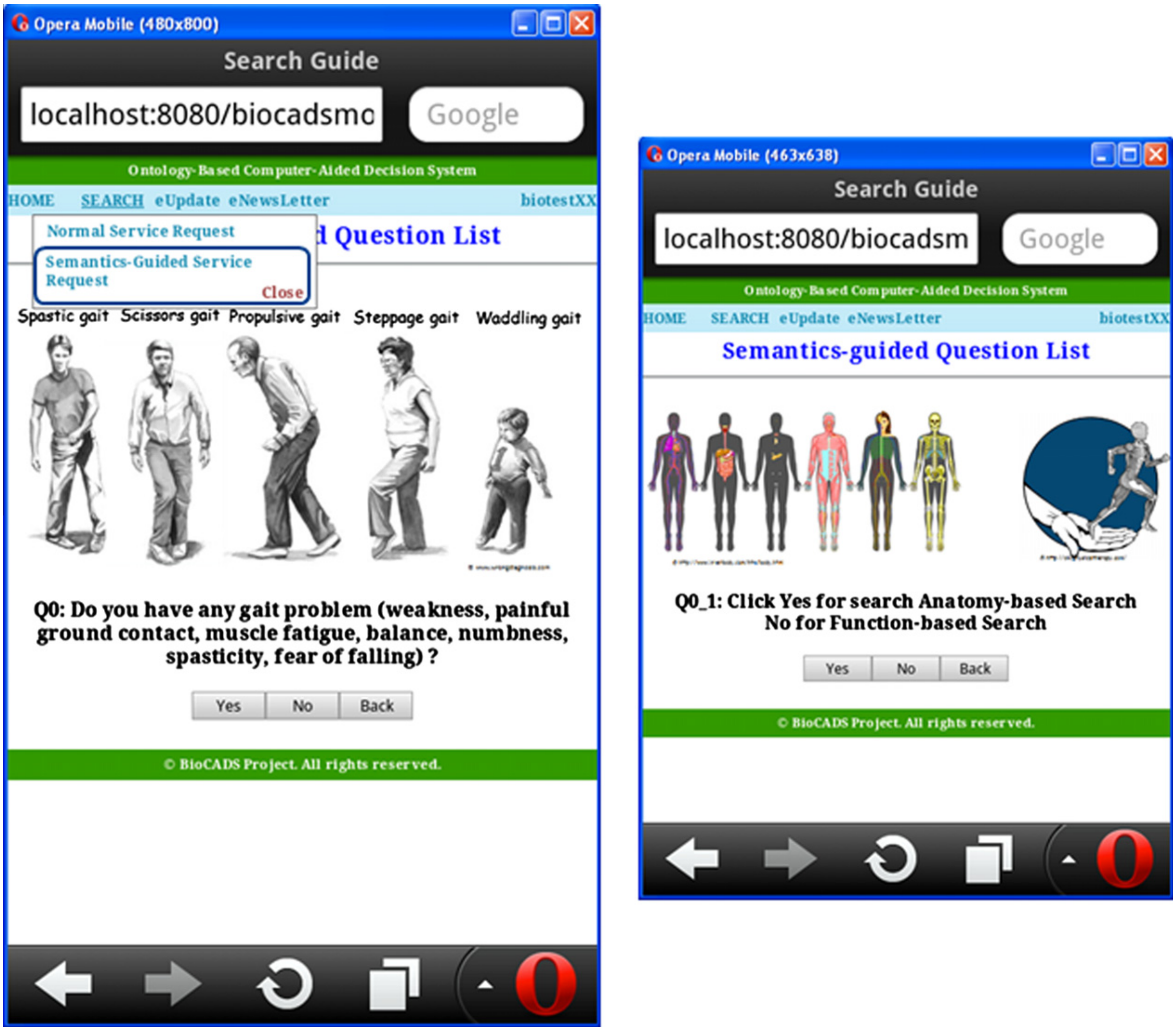

Fig. 10. Web-based interfaces for the mobile device of the semantics-guided service request.

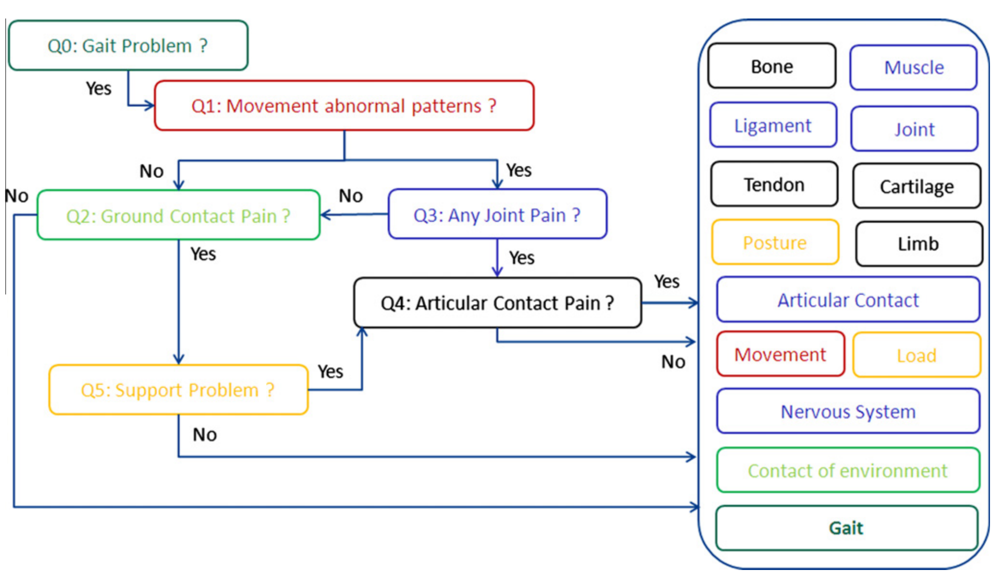

Fig. 11. Activation principles of target concepts based on OSMMI ontology and biomechanics knowledge.

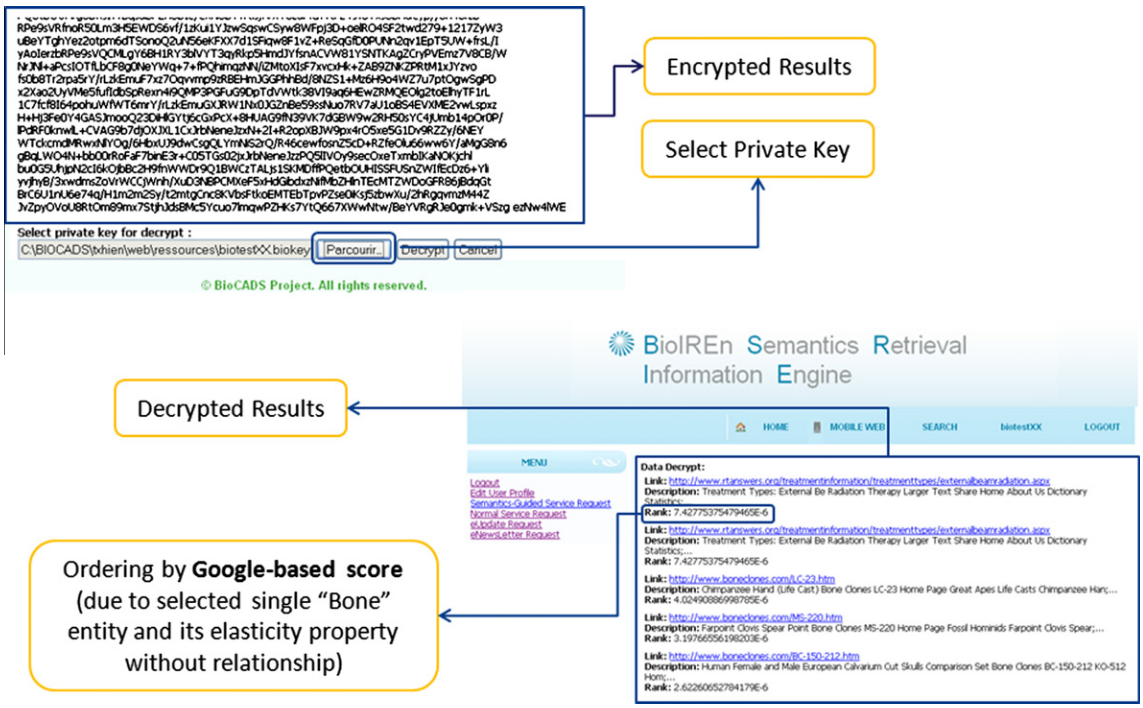

Fig. 12. Retrieved results ordered by using Google-based PageRank algorithm due to the selected unique concept "Bone".

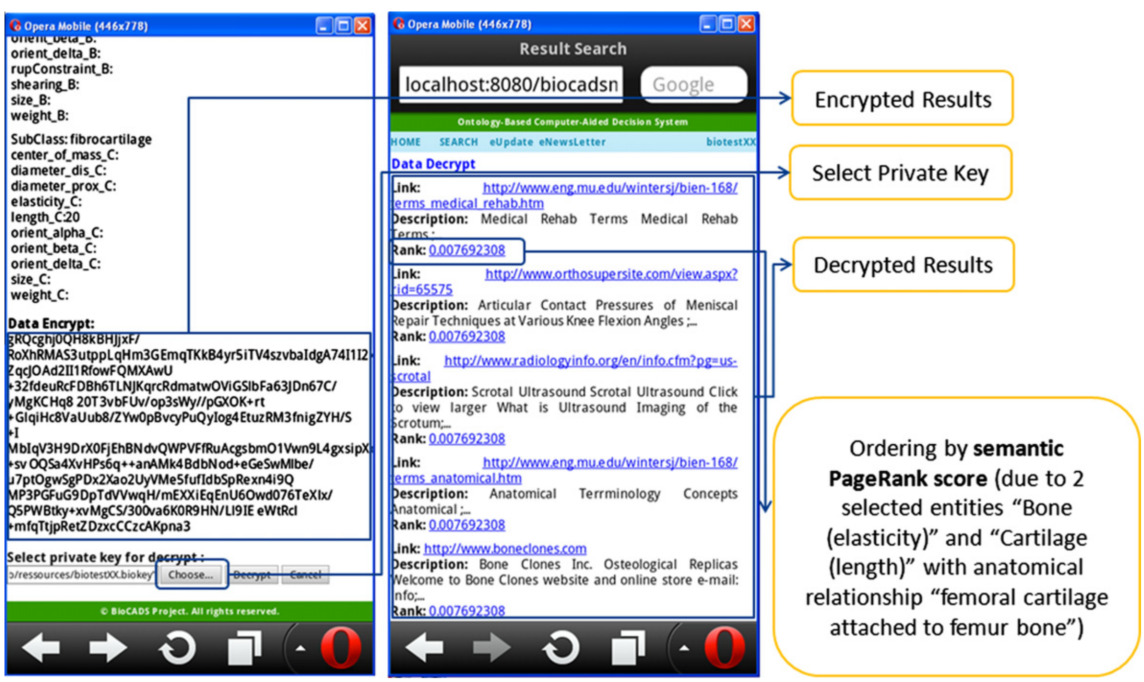

Fig. 13. Retrieved results ordered by using Semantics-based PageRank algorithm due to 2 selected concepts such as "Bone" (with elasticity as selected property) and "Cartilage" (with length as selected property).

keyword is put into the HMSR context) and other information without meaning in related context (e.g. "pain" keyword can be used with French language meaning (i.e. "bread" meaning) are filtered by the annotation process and semantic PageRank algorithm.

A detailed summary of other testing case queries are presented in Table 1. The number of relevant retrieved results is computed based on the Google and our PageRank scores. For example, in the first case with two concepts "Bone" and "Muscle", based on our semantics-based PageRank score, the end user obtains more relevant and precise results. In the second case with many concepts and properties, only semantics-based PageRank algorithm provides relevant results. In the third case with single concept, Google PageRank algorithm provides relevant results due to the lack of semantic relationships in search request query.

A performance comparison between Google engine and our knowledge-based search engine showed that our semantic search

Table 1
A testing case report of our knowledge-based search engine.

Testing case query	Google PageRank		Semantic PageRank	
	Retrieved time (s)	Relevant retrieved results	Retrieved time (s)	Relevant retrieved results
Bone: cacaneus_foot (length), Muscle: Pyramidal (force), Tendon: Gastrocnemius (length)	15	674 websites	5	6310 websites
Cartilage_cartilage (force), fibrocartilage (elasticity) foot_floor: (force), coxo_femoral (force), normal_gait (ankleTorque), anterior_tibio_fibular (elasticity), rotation (velocity) quadriceps (length)	15	0 websites	5	3039 websites
Muscle: quadriceps (length)	5	2605 websites	5	0 websites

Table 2
A performance evaluation report of our knowledge-based search engine.

Query	Query content	Google engine P@100[a]	Our knowledge-based search engine P@100[a]
Q#1	Joint: knee (force)	0.84	1
Q#2	Tendon: Achilles (length)	0.9	1
Q#3	Cartilage: hyaline (length)	0.51	1
Q#4	Muscle: quadriceps (length)	0.72	1
Q#5	Movement: rotation (sagittal)	0.94	1

[a] $P@100 = \frac{a}{a+c}$: a is the number of relevant retrieved websites from the 100-first retrieved records; c (=100 − a) is the number of irrelevant retrieved websites from the 100-first retrieved records.

engine reaches a precision of 1 in many case (Table 2). Note that this comparison was performed only on the 100-first retrieved records.

All testing cases were performed on a content-based database including 17859 relevant website models, a Tomcat local web server (Intel Xeon 2.67 GHz, RAM 3 GB) using Windows XP as operating system. The time of Google PageRank score computing is about 25 min.

3.6. Comparative functional analysis

The comparative report between our knowledge-based search engine, a conventional search engine (Google) and a semantic search engine[2] [45] is presented in Table 3. By using the OSMMI ontology, our search query set up is based on a user-interest and domain-based approach in comparison to the keywords-based approach of Google engine and tuple-based table approach of SIREn system. Our ordering results are based on the semantics terminologies and their anatomical and structural relationships while Google and SIREn use link structure-based principle. Moreover, our search engine provide a persistent level for the query set up as well as for the result storage while queries and results from Google and SIREn systems are volatile. Furthermore, our search engine provides an additional information synthesis service and security level according to other search engines.

4. Discussion

Our knowledge-based personalized search engine integrated all necessary components to provide a technological solution for the need of more accurate and reliable human musculoskeletal resources related to learning (education and training) and medical processes (diagnosis, treatment, monitoring), pathological knowledge and expertise. Our search engine was developed using a three-semantics-level approach. The first semantic level is built at the query set up by using our available ontology giving a user-interest and domain-oriented structure. This facilitates the query task set up and increases the semantics of the query contents [14,46–48]. The second semantic level relates to the semantic annotation of crawled website models. Based on the OSMMI ontology, only semantic-significant website models were annotated, stored and reused. This elite-based strategy allows only the more accurate and reliable HMSR information to be used in our semantic search engine. The final semantic level deals with the result ordering. A semantics-based PageRank score was defined to obtain the most suitable results for a specific query. In fact, our computational musculoskeletal OSMMI ontology [2] provides fundamental semantic structure for the query set up, the webpage annotation and the semantics PageRank score computing to have more precise query set up as well as more semantics-rich query results [15] leading to more accurate and reliable HMSR information.

Current web-based information retrieval systems such as Google or Yahoo engines used keyword-based approaches leading to the ambiguation of query set up for a specific request. Moreover, the ranking of the retrieved results is based on link-in and link-out principle (i.e. random surfing model) leading to the inappropriate and unsatisfactory retrieved results [49]. To tackle this problem, some studies adapted or extended the standard PageRank principle to develop more specific ranking algorithms such as ranking based on the sensitivity of the topic [50] or the link-based approach [51]. Semantic disambiguation of text word sense was also performed using logical inferences coupled with a PageRank-style algorithm [52]. Furthermore, semantic search engines integrating knowledge-based approaches such as relation-based Page approach [53] or the semantic ranking based on the heterogeneity of relationships between resources [38] were also developed. The first main difference between our semantic PageRank algorithm and these approaches is the integration of query content in the computing of our semantic RageRank score. Thus the retrieved results are assessed and ranked according to the specific content of the query involved. The second difference is the use of a content-based database to compute the semantic PageRank score. Consequently, the semantic annotation process plays a semantic filtering role to avoid inappropriate ranked web sites models in the HMSR context. The third difference is the weighted level-dependent character of our dedicated first semantic rule regarding the semantic importance of property according to an entity in the HMRS context.

The use of semantic web services is the originality of our search engine. Three innovative semantics web services such as semantic request service, e-Update service and e-Newsletter service were developed for the HMSR information in the Biomechanics field. This provides content-based request set up according to the user's needs and user's contents through the *Template and User Service Ontologies*, the conceptual transcription, the exchange and the interoperability capacities by using XML technology [2,11,23]. Based on these semantic structures, a personalized search engine was developed giving the personalized search query set up as well as the deeper annotation of crawled webpage models. Moreover, advanced ordering mechanism was developed. Consequently, this can be combined with the Google PageRank algorithm in a knowledge-based search engine. In fact, all these components allow the

[2] http://siren.sindice.com/.

Table 3
Comparative report between our knowledge-based search engine, a conventional search engine (Google) and a semantic search engine.

Criterion	Our BioIREn®	Google	SIREn
Type	Semantic IR System	Traditional IR system	Semantic IR System
Query set up	User-interest and domain-based (ontology)	Keywords	Tuple Table Query Model
Ordering results	Semantics-based terminologies and relationship (anatomical and structural)	Input and output links	Weighted PageRank (Hierarchical Link Analysis)
Query storage	Persistent database	Volatile	Volatile
Results storage	Persistent database	Volatile	Volatile
Information synthesis	eUpdate and eNewsletter	Unavailable	Unavailable
Information security	Cryptography mechanism	Insecurity/Unavailable	Insecurity/Unavailable
Interfaces	Web and Mobile	Web and Mobile	Web
Annotation, indexing (Crawler)	Metadata => Ontology	Metadata => Keywords	N-Triples
Multimedia contents	Text-structured input	Text, image input	Text-structured input
Search Language	English	Multiple languages (en, fr, ita, etc.)	–

accurate and reliable responses according to a specific personalized request are dynamically obtained and updated [42,43,54,55].

A cryptography algorithm was applied to provide a security-based authentication mechanism for our web-based search engine. This allows the security and the integrity of medical information to be protected and warranted [56]. Moreover, the development of our crawler is based on a multi-threads multi-agent approach. This parallel mechanism allows a significant search database to be acquired in a shorter time. In particular, a persistent database was created to keep the information alive which can be updated dynamically. Furthermore, a semantics-rich annotation mechanism was applied on each crawled webpage model from the Internet to create a content-based database [23].

In addition to the normal search, a semantics-guided intelligent mechanism was proposed to refine the requested *User Service Ontology*. This questionnaire-based guide allows the end user to perform a semantic-rich query set up with more detailed request information according to his (her) specific need. Moreover, the choice of web-based interfaces for personal computers and mobile devices shows that our search engine is independent regarding the operating systems and the services provider [57–59]. This allows the end user to connect remotely into semantic web services via a web browser. In fact, these remote services with user-friendly usage allow us to reduce the medical cost and resources.

Comparison with other search engines shows that our search system has additional useful functions such as the use of a user-interest and domain-oriented ontology for query set up or persistent strategy for query and result storage purposes. In addition to the information synthesis capacity, a security-based mechanism also shows its useful application for a medical information retrieval system. In particular, a semantic-rich approach was applied for annotating, indexing and ordering the crawled and retrieved website models leading to more precise and detailed query/retrieval information [11,14,23,29,35,38]. However, our search engine is dedicated for the English language and text-structured contents while Google provides multi-languages and multimedia search capacities.

Regarding the evaluation of our semantic search engine, retrieved records were considered as relevant records thanks to the use of context-based annotation process and semantic PageRank algorithm leading to the disambiguation of the content on the website models. Thus, our semantic search system reaches a precision of 1 in many cases which are similar with those reported in Fazzinga et al. [11] for their semantic search system. However, current systematic performance evaluation of a semantic web search is not obvious due to the lack of a more objective and comprehensive approach regarding the evaluation of the relevant and irrelevant records retrieved in the semantic context. Especially, the evaluation of relevant records not retrieved is not straightforward. New efforts will be investigated to develop new criteria and metrics for a more comprehensive systematic evaluation approach of the semantic search tools, as the objectives of the European SEALS (Semantic Evaluation At Large Scale) project (www.seals-project.eu).

One of the limitations of our knowledge-based search engine relates to our musculoskeletal ontology. Our first prototype is only dedicated for the musculoskeletal system of the human lower limbs and related orthopedic pediatric disorders. However, our methodology could be extrapolated to other human systems (e.g. cardiovascular or nervous systems) and disorders (e.g. low back pain or arthritis) to develop a full version of our semantic search engine.

Another limitation deals with the computing of the semantic PageRank score for each crawled website model. This process requires intensive computer resources in relation to the memory and the CPU time. Consequently, more powerful and dedicated database server (center) are needed for the deployment of our semantic search engine in real usage context.

In conclusions, a knowledge-base personalized search engine was developed to access the Web-based Human Musculoskeletal System Resource (HMSR) information in the Biomechanics field. Our knowledge-based search engine was based on the principle of semantic web services to acquire dynamically accurate and reliable HMSR information by a semantic processing and visualization approach. A security-enhanced mechanism was applied to protect the medical information. A semantic-based PageRank score was also defined and implemented to propose a semantics-rich score dedicated for our semantic search engine in the Biomechanics field. All these components allow different users such as orthopedic patients and experts or healthcare system managers or medical students to access remotely into useful, accurate, reliable and good-quality HMSR information for their learning and medical purposes.

Conflict of Interest

The authors declare that there is no conflict of interest dealing with this present work.

Acknowledgments

The authors would like to thank the anonymous reviewers for their comments leading to improve this paper. The present work was funded by the Picardie Region.

References

[1] Chapple A, Campbell S, Rogers A, Roland M. Users' understanding of medical knowledge in general practice. Soc Sci Med 2002;54(8):1215–24.

[2] Dao TT, Marin F, Ho Ba Tho MC. Ontology of the musculoskeletal system of lowers limbs. In: Proceedings of the 29th annual international conference of the IEEE engineering in medicine and biology society; 2007. p. 386–9.
[3] Armoni A. Knowledge acquisition for medical diagnosis systems. Knowledge-Based Syst 1995;8(4):223–6.
[4] Dao TT, Marin F, Ho Ba Tho MC. Ontology-based computer-aided decision system: a new architecture and application concerning the musculoseletal system of the lower limbs. In: Proceedings of the 4th European congress of the international federation for medical and biological engineering. Berlin, Heidelberg: Ed Springer; 2008. p. 1540–3.
[5] Dao TT, Marin F, Bensahel H, Ho Ba Tho MC. Computer-aided decision system applied to the clubfeet deformities. Adv Exp Med Biol 2011;696(7):623–35.
[6] Aanestad M, Jensen TB. Building nation-wide information infrastructures in healthcare through modular implementation strategies. J Strateg Inf Syst 2011;20(2):161–76.
[7] Carlile S, Barnet S, Sefton A, Uther J. Medical problem based learning supported by intranet technology: a natural student centred approach. Int J Med Inform 1998;50(1–3):225–33.
[8] Bernardo V, Ramos MP, Plapler H, de Figueiredo LFP, Nader HB, Ancão MS, et al. Web-based learning in undergraduate medical education: development and assessment of an online course on experimental surgery. Int J Med Inform 2004;73(9–10):731–42.
[9] Elder WG, Dassow PL, Bruckner GG, Stratton TD. Impact of two types of Internet-based information on medical students' performance in an Objective Structured Clinical Exam (OSCE). Patient Educ Couns 2009;76(1):5–9.
[10] Lee S, Kim K. Factors affecting the implementation success of Internet-based information systems. Comput Hum Behav 2007;23(4):1853–80.
[11] Fazzinga B, Gianforme G, Gottlob G, Lukasiewicz T. Semantic Web search based on ontological conjunctive queries. Web Semant: Sci Serv Agents World Wide Web 2011;9(4):453–73.
[12] Koning R, Grosso P, de Laat C. Using ontologies for resource description in the CineGrid exchange. Future Generation Comput Syst 2011;27(7):960–5.
[13] Beckett D. The design and implementation of the Redland RDF application framework. Comput Netw 2002;39(5):577–88.
[14] Fernández M, Cantador I, López V, Vallet D, Castells P, Motta E. Semantically enhanced information retrieval: an ontology-based approach. Web Semant: Sci Serv Agents World Wide Web 2011;9(4):434–52.
[15] Yang SY. An ontological website models-supported search agent for web services. Expert Syst Appl 2008;35:2056–73.
[16] Yang SY. OntoCrawler: a focused crawler with ontology-supported website models for information agents. Expert Syst Appl 2010;37:5381–9.
[17] Niaraki AS, Kim K. Ontology based personalized route planning system using a multi-criteria decision making approach. Expert Syst Appl 2009;36:2250–9.
[18] Denker G, Kagal L, Finin T. Security in the Semantic Web using OWL. Inf Secur Tech Rep 2005;10:51–8.
[19] Hendler J, Golbeck J. Metcalfe's law, Web 2.0, and the Semantic Web. Web Semant: Sci Serv Agents World Wide Web 2008;6:14–20.
[20] Ortiz G, Prado AGD. Improving device-aware Web services and their mobile clients through an aspect-oriented, model-driven approach. Inf Softw Technol 2010;52:1080–93.
[21] Song K, Lee KH. Generating multimodal user interfaces for Web services. Interact Comput 2008;20:480–90.
[22] Toninelli A, Corradi A, Montanari R. Semantic-based discovery to support mobile context-aware service access. Comput Commun 2008;31:935–49.
[23] Agarwal S, Handschuh S, Staab S. Annotation, composition and invocation of semantic web services. Web Semant: Sci Serv Agents World Wide Web 2004;2:31–48.
[24] Puttonen J, Lobov A, Soto MAC, Lastra JLM. A semantic web services-based approach for production systems control. Adv Eng Inform 2010;24:285–99.
[25] García-Sánchez F, Fernández-Breis JT, Valencia-García R, Gómez JM, Martínez-Béjar R. Combining Semantic Web technologies with Multi-Agent Systems for integrated access to biological resources. J Biomed Inform 2008;41:848–59.
[26] Gibbins N, Harris S, Shadbolt N. Agent-based semantic web services. Web Semant: Sci Serv Agents World Wide Web 2004;1:141–54.
[27] Izza S, Vincent L, Burlat P. Exploiting semantic web services in achieving flexible application integration in the microelectronics field. Comput Ind 2008;59:722–40.
[28] Lupiani-Ruiz E, García-Manotas I, Valencia-García R, García-Sánchez F, Castellanos-Nieves D, Fernández-Breis JT, et al. Financial news semantic search engine. Expert Syst Appl 2011;38(12):15565–72.
[29] Weiand K, Kneißl F, Łobacz W, Furche T, Bry F. Pest: fast approximate keyword search in semantic data using eigenvector-based term propagation. Inf Syst 2012;37(4):372–90.
[30] Hamdi MS. SOMSE: a semantic map based meta-search engine for the purpose of web information customization. Appl Soft Comput 2011;11(1):1310–21.
[31] Batzios A, Mitkas PA. WebOWL: a semantic web search engine development experiment. Expert Syst Appl 2012;39(5):5052–60.
[32] Halpin H, Lavrenko V. Relevance feedback between hypertext and Semantic Web search: frameworks and evaluation. Web Semant: Sci Serv Agents World Wide Web 2011;9(4):474–89.
[33] Sauer UG, Kneuer C, Tentschert J, Wächter T, Schroeder M, Butzke D, et al. A knowledge-based search engine to navigate the information thicket of nanotoxicology. Regul Toxicol Pharmacol 2011;59(1):47–52.
[34] Kim C, Park S, Kwon K, Chang W. An empirical study of the structure of relevant keywords in a search engine using the minimum spanning tree. Expert Syst Appl 2012;39(4):4432–43.
[35] Herskovic JR, Cohen T, Subramanian D, Iyengar MS, Smith JW, Bernstam EV. MEDRank: using graph-based concept ranking to index biomedical texts. Int J Med Inform 2011;80(6):431–41.
[36] Han L, Sun L, Chen G, Xie L. ADSS: an approach to determining semantic similarity. Adv Eng Softw 2006;37(2):129–32.
[37] Nakatsuji M, Yoshida M, Ishida T. Detecting innovative topics based on user-interest ontology. Web Semant: Sci Serv Agents World Wide Web 2009;7:107–20.
[38] Ning X, Jin H, Wu H. RSS: a framework enabling ranked search on the semantic web. Inf Process Manage 2008;44(2):893–909.
[39] Mecca G, Raunich S, Pappalardo A. A new algorithm for clustering search results. Data Knowledge Eng 2007;62(3):504–22.
[40] Huang YM, Kuo YH, Lin YT, Cheng SC. Toward interactive mobile synchronous learning environment with context-awareness service. Comput Educ 2008;51:1205–26.
[41] Becker C, Bizer C. Exploring the geospatial semantic web with DBpedia mobile. Web Semant: Sci Serv Agents World Wide Web 2009;7:278–86.
[42] Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web. Stanford InfoLab. Technical Report; 1999. p. 1–17.
[43] Sun H, Wei Y. A note on the PageRank algorithm. Appl Math Comput 2006;179:799–806.
[44] Lin Y, Shi X, Wei Y. On computing PageRank via lumping the Google matrix. J Comput Appl Math 2009;224:702–8.
[45] Delbrua R, Campinasa S, Tummarello G. Searching web data: an entity retrieval and high-performance indexing model. Web Semant: Sci Serv Agents World Wide Web 2012;10:33–58.
[46] Chiang RHL, Chua CEH, Storey VC. A smart web query method for semantic retrieval of web data. Data Knowledge Eng 2001;38(1):63–84.
[47] Conesa J, Storey VC, Sugumaran V. Improving web-query processing through semantic knowledge. Data Knowledge Eng 2008;66(1):18–34.
[48] Bianchinia D, de Antonellisa V, Pernicib B, Plebani P. Ontology-based methodology for e-service discovery. Inf Syst 2006;31:361–80.
[49] Ding L, Pan R, Finin T, Joshi A, Peng Y, Kolari P. Finding and ranking knowledge on the semantic web. In: Proceedings of the 4th international semantic web conference; 2005. p. 156–70.
[50] Haveliwala TH. Topic-sensitive pagerank. In: WWW'02; 2002. p. 517–26.
[51] Baeza-Yates R, Davis E. Web page ranking using link attributes. In: WWW'04; 2004. p. 328–9.
[52] Mihalcea R, Tarau P, Figa E. PageRank on semantic networks, with application to word sense disambiguation. In: COLING '04 proceedings of the 20th international conference on computational linguistics; 2004. p. 1–7.
[53] Lamberti F, Sanna A, Demartini C. A relation-based page rank algorithm for semantic web search engines. IEEE Trans Knowledge Data Eng 2009;21(1):123–36.
[54] Wei W, Barnaghi P, Bargiela A. Rational research model for ranking semantic entities. Inf Sci 2011;181:2823–40.
[55] Liu M, Shen W, Hao Q, Yan J. An weighted ontology-based semantic similarity algorithm for web service. Expert Syst Appl 2009;36:12480–90.
[56] Antonopoulou E, Karetsos ST, Maliappis M, Sideridis AB. Web and mobile technologies in a prototype DSS for major field crops. Comput Electron Agric 2010;70:292–301.
[57] Motiwalla LF. Mobile learning: a framework and evaluation. Comput Educ 2007;49:581–96.
[58] Garrett BM, Jackson C. A mobile clinical e-portfolio for nursing and medical students, using wireless personal digital assistants (PDAs). Nurse Educ Today 2006;26:647–54.
[59] Madeiraa RN, Sousaa JL, Piresa VF, Estevesa L, Diasb OP. A mobile and web-based student learning system. Proc Social Behav Sci 2009;1:2441–8.