Computer Engineering and Intelligent Systems www.iiste.org
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.2, 2012

Integrated Approach for Domain Dimensional Information Retrieval System by Using Neural Networks

R. Kamalakar, M.Anjan Kumar, P.Pradeep Kumar, P.Sai Prasad

*kamalsee@gmail.com, anjanind@gmail.com , pkpuram@yahoo.com , saiprsadcse@gmail.com

Abstract:

Search engines are the most commonly used type of tool for finding relevant information on the Internet. However, today's search engines are far from perfect. Typical search queries are short, often one or two words, and can be ambiguous therefore returning inappropriate results. A precise search engine adapted to professional environments which are characterized by a domain (e.g. medicine, law, sport, and so on). In our approach, each domain has its own terminology (i.e. a set of terms that denote its concepts: team, player, etc.) and it is organized along dimensions, such as person, location, etc. The research work is focuses on personalization of information retrieval systems to achieve this we require one architecture that is to developed with immense ground knowledge on open source technologies and great Neural networks and Information retrieval system and there scope and existence. The architecture termed as ISA integrated service architecture.

1. Introduction

1.1 Artificial Neural Networks

The term artificial intelligence was first coined in 1956, at the Dartmouth conference, and since then Artificial Intelligence has expanded because of the theories and principles developed by its dedicated researchers. Through its short modern history, advancement in the fields of AI have been slower than first estimated, progress continues to be made. From its birth 4 decades ago, there have been a variety of AI programs, and they have impacted other technological advancements

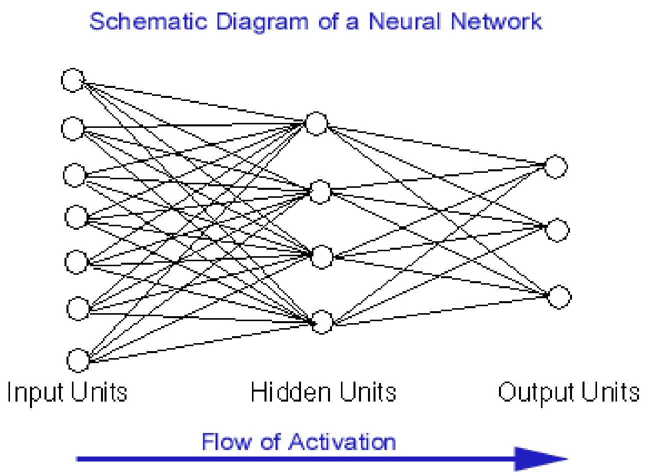
A neural network is, in essence, an attempt to simulate the brain. Neural network theory revolves around the idea that certain key properties of biological neurons can be extracted and applied to simulations, thus creating a simulated (and very much simplified) brain. The first important thing to understand then, is that the components of an artificial neural network are an attempt to recreate the computing potential of the brain. The second important thing to understand, however, is that no one has ever claimed to simulate anything as complex as an actual brain. Whereas the human brain is estimated to have something on the order of ten to a hundred billion neurons, a typical artificial neural network (ANN) is not likely to have more than 1,000 artificial neurons. Artificial Intelligence is a combination of computer science, physiology, and philosophy.

Before discussing the specifics of artificial neural nets though, let us examine what makes real neural nets - brains - function the way they do. Perhaps the single most important concept in neural net research is the idea of connection strength. Neuroscience has given us good evidence for the idea that connection strengths - that is, how strongly one neuron influences those neurons connected to it - are the real information holders in the brain. Learning, repetition of a task, even exposure to a new or continuing stimulus can cause the brain's connection strengths to change, some synaptic connections becoming reinforced and new ones are being created, others weakening or in some cases disappearing altogether.

A neuron may sum its inputs, or average them, or something entirely more complicated. Each of these behaviors can be represented mathematically, and that representation is called the transfer function. It is often convenient to forget the transfer function, and think of the neurons as being simple addition machines, more activity in equals more activity out. This is not really accurate though, and to develop a good understanding of an artificial neural network, the transfer function must be taken into account. A Armed with these three concepts: Connection, strength Inhibition/ Excitation , and the Transfer Function, we can now look at how artificial neural nets are constructed. In theory, an artificial neuron (often called a 'node') captures all the important elements of a biological one. Nodes are connected to each other and the strength of that connection is normally given a numeric value between -1.0 for maximum inhibition, to +1.0 for maximum excitation. All values between the two are acceptable, with higher magnitude values indicating stronger connection strength. The transfer function in artificial neurons whether in a computer simulation, or actual microchips wired together, is typically built right into the nodes' design. Perhaps the most significant difference between artificial and biological neural nets is their organization. While many types of artificial neural nets exist, most are organized according to the same basic structure (see diagram). There are three components to this organization: a set of input nodes, one or more layers of 'hidden' nodes, and a set of output nodes. The input nodes take in information, and are akin to sensory organs. Whether the information is in the form of a digitized picture, or a series of stock values, or just about any other form that can be numerically expressed, this is where the net gets its initial data. The information is supplied as activation values, that is, each node is given a number, higher numbers representing greater activation. This is just like human neurons except that rather than conveying their activation level by firing more frequently, as biological neurons do, artificial neurons indicate activation by passing this activation value to connected nodes. After receiving this initial activation, information is then passed through the network. Connection strengths, inhibition/excitation conditions, and transfer functions determine how much of the activation value is passed on to the next node. Each node sums the activation values it receives, arrives at its own activation value, and then passes that along to the next nodes in the network (after modifying its activation level according to its transfer function). Thus the activation flows through the net in one direction, from input nodes, through the hidden layers, until eventually the output nodes are activated. If a network is properly trained, this output should reflect the input in some meaningful way. For instance, a gender recognition net might be presented with a picture of a man or woman at its input nodes and must set an output node to 0.0 if the picture depicts a man, or 1.0 for a woman. In this way, the network communicates its knowledge to the outside world.

1.2 Information retrieval system

INFORMATION Retrieval Systems (IRS) are nowadays very popular, mainly due to the popularity of the Web. Most IRS on the Web (also called search engines) are not really domain-oriented: the same techniques are used to index any document. We think that there is a niche for domain-specific IRS



+

2. Relevance Analysis

once the document domain is known, certain assumptions can be made and specific knowledge can be used. Users are then allowed to utilize much more precise queries than the usual small set of keywords in use for Web search engines. In professional environments, IRS should be able to process precise queries, mostly due to its use of a specific terminology, but also because the retrieved information is meant to be part of a user task (diagnose a disease, write a report, etc.).

In professional environments, there is also a growing need for accessing information about specific domain documents in many languages and many types of media. In this paper, we present a precise search engine adapted to professional environments that are characterized by a domain (e.g. Medicine, Law, Sport, and so on). In our approach, each domain has its own terminology (i.e. a set of terms that denote its concepts) and it is organized along dimensions, such as *Person*, *Location*, etc.

Dimensions, as described below, are defined by concepts and semantic relationships that represent a particular perspective or point of view on the corresponding domain. We mainly use the notion of domain dimension to *i*) precisely index document content and *ii*) implement an interactive interface that allows users to precisely describe his or her information need, and therefore precisely access a document collection. Our main goal through this system is to allow users fluid access to a digital library that contains documents belonging to specific domains, written in different languages, and using differnt medias. In particular, our system provides the user at all times with a feeling of control and understanding. It therefore provides a keyword search combined with a flexible navigation system. This combination allows a user to select a domain of his interest, build his query, expand and refine it, and select the language and the medias of the search results**.**

3. ***Domain Dimensions***

We use domain dimensions for solving domain-specific precise queries that are characterized by a specialized terminology and a complex semantic structure. In this case, domain dimensions are used to extract the specialized vocabulary and therefore highlight the relevant elements that contribute to the description of a document (or query) semantic content. For example, through our dimension-based model, a journalist wishing to write a newspaper article can formulate his query as follows: “Give me documents dealing with the French General who created the security zone during the Balkans conflict”. Our system is able to recognize domain dimensions and use them to precisely answer this query: *Person* (French General), *Location* (Balkans, security zone), *Event* (Balkans conflict).

4. Search Analysis

The main idea behind a search interface based on domain dimensions is quite simple. Rather than creating one large category hierarchy, we build a set of category hierarchies, each of which corresponds to a different dimension (facet) relevant to the domain described in the collection to be navigated. This representation is also known as hierarchical faceted categories. In dimension-based search interfaces, each domain has a set of dimensions and each dimension has a hierarchy of concepts. After the dimensions’ hierarchies are designed, each document in the collection can be assigned to many concepts from the hierarchies. For example, in the medical domain, the dimension hierarchies can include *Human Anatomy* (Head, Brain, Femur, etc. with *part-of* relationships), *Pathology* (Cancer, fracture, lesion, etc., with is-a relationships), *Image Modality* (MRI, x-ray, ultrasound, etc., with is-a relationships) and so on. Thus, an *MRI* image describing a *fracture* of a *femur* might be assigned to (indexed by):

Anatomy > Musculoskeletal System > Skeleton > Bone and

Bones > Bones of Lower Extremity > Foot Bones > Leg Bones> Femur

Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.2, 2012

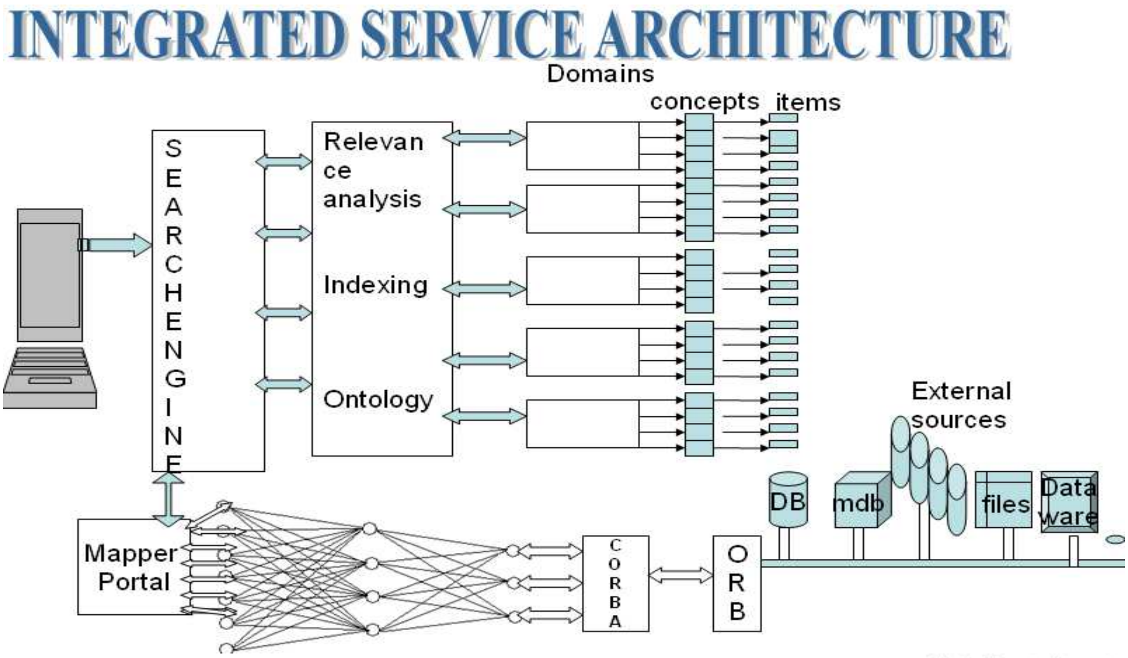
Pathology > Disorders of Environmental Origin > Wounds

and Injuries > Fractures, Bone > Femoral Fractures

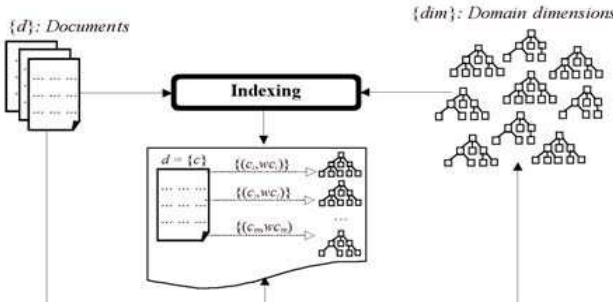
Modality > Diagnostic Techniques and Procedures > Diagnostic

Imaging > Magnetic Resonance Imaging

When a concept within a dimension hierarchy is selected within the interface, all documents that have been assigned to that concept are retrieved (and displayed). When concepts from different hierarchies are selected, the system builds a query that is a conjunct of disjuncts over the selected concepts and their sub concepts.



5. Document Indexing



The architecture of our system is presented in Figure 3.There are mainly three steps in our system design: *i*) defining the domain dimensions relevant to the given document collection; *ii*) multidimensional document indexing, which matches associate documents to the corresponding domain dimensions;

A. External Resources-Based Domain Dimension Definition

Our aim is to access a multimedia multilingual document collection through a dimensions-based search interface. Instead of using a specific indexing technique for each language and each media, we propose to use a unique technique for all documents independent of their languages or their media type. This technique is based on conceptual indexing that consists in representing documents (queries) by concepts instead of ambiguous descriptors (words extracted from text, features extracted from audiovisual information such as colour, shape, texture, motion, audio frequency, etc

6.INTEGRATED SERVICE ARCHITECTURE

An ISA integrated service architecture is proposed to do the efficient information retrieval form multiple heterogeneous data sources. The efficency can be achieved by using the integrated technologies of Neural Networks and Object Resource Broker through CORBA.

The ISA architecture describes that the user can connect the search engine with own local resources to remote and search engine and there functionalities and comes under the domain of IRS. In previous sections we discussed the concept on relevance analysis and indexing with multiple domain dimensionalities .

Neural networks are the decision making tools and forwarding and back warding approaches when the desired date are to be fetched from the source by any neuron it will immediately sends to the MAPPER PORTAL . portal is a third party tool to integrate both environment for functional processing .

Why should I use corba and orbs?

The only the strait forward solution is that neutrability mean is an open source. By using orb's large immense features we can connect any type of data source like relational, object based, spatial, time series , file system and even data ware houses also. Due to implementation cost reduction and flexibility uses the orb.

The architecture implementation taken place and some important portion of integrations are carried on my work after success full implementation we will produce the results and dataset and constraint elaborately.

7. CONCLUSION

The research work is focuses on personalization of information retrieval systems to achieve this we require one architecture that is to developed with immense ground knowledge on open source technologies and great Neural networks and Information retrieval system and there scope and existence.

The architecture implementation taken place and some important portion of integrations are carried on my work after success full implementation we will produce the results and dataset and constraint elaborately.

8. References

Flexible building of specialized information retrieval systems application to the management of APL functions as a programming aid. Authors: Jean-Claude pages,Anne Mauboussin

Contemporary model and initiatives for information evalution. Systems: M.Anjan kumar, R.kamalakar. *Proceedings of the 63rd Internationalconference AMCM.2010*

"Neural networks with advanced bonding sectors ,"by M.ANJANKUMAR , R.KAMALAKAR .A.HUSSAIN in *Proceedings of the 3rdNationl Conference inf VITS(n6),knr Through ISTE CHAPTER ,2011*7.

Yee, K.P., Swearingen, K., Li, K., Hearst, M., "Faceted metadata for image search and browsing," in *CHI '03: Proceedings of the conference*
on Human factors in computing systems, ACM Press, pp. 401-408, 2003.

M¨akel¨a, E., Hyv¨onen, E., Saarela, S., "Ontogator - a semantic view based search engine service for web applications," in *International Semantic Web Conference*, pp. 847-860, 2006.

M¨akel¨a, E., Hyv¨onen, E., Sidoroff, T., "View-based user interfaces for information retrieval on the semantic web," in *ISWC-2005 WorkshopEnd User Semantic Web Interaction*, November 2005.

Sacco, G.M., "Research results in dynamic taxonomy and faceted searchsystems," in *DEXA Workshops*, IEEE Computer Society