INFORMATION DISCOVERY IN DISTRIBUTED SYSTEMS – A NOVEL CONCEPT BASED ON MODELLING OF INFORMATION ELEMENTS

ABSTRACT

Sharing of information with those in need of it has always been an idealistic goal of networked environments. With the proliferation of computer networks, information is so widely distributed among systems, that it is imperative to have well-organized schemes for retrieval and also discovery. This thesis attempts to investigate the problems associated with such schemes and suggests a software architecture, which is aimed towards achieving a meaningful discovery. Usage of information elements as a modelling base for efficient information discovery in distributed systems is demonstrated with the aid of a novel conceptual entity called infotron.

The investigations are focused on distributed systems and their associated problems. The study was directed towards identifying suitable software architecture and incorporating the same in an environment where information growth is phenomenal and a proper mechanism for carrying out information discovery becomes feasible. An empirical study undertaken with the aid of an election database of constituencies distributed geographically, provided the insights required. This is manifested in the Election Counting and Reporting Software (ECRS) System. ECRS system is a software system, which is essentially distributed in nature designed to prepare reports to district
administrators about the election counting process and to generate other miscellaneous statutory reports.

Most of the distributed systems of the nature of ECRS normally will possess a "fragile architecture" which would make them amenable to collapse, with the occurrence of minor faults. This is resolved with the help of the penta-tier architecture proposed, that contained five different technologies at different tiers of the architecture.

The results of experiment conducted and its analysis show that such an architecture would help to maintain different components of the software intact in an impermeable manner from any internal or external faults. The architecture thus evolved needed a mechanism to support information processing and discovery. This necessitated the introduction of the novel concept of infotrons. Further, when a computing machine has to perform any meaningful extraction of information, it is guided by what is termed an infotron dictionary.

The other empirical study was to find out which of the two prominent markup languages namely HTML and XML, is best suited for the incorporation of infotrons. A comparative study of 200 documents in HTML and XML was undertaken. The result was in favor of XML.

The concept of infotron and that of infotron dictionary, which were developed, was applied to implement an Information Discovery System (IDS). IDS is essentially, a system, that starts with the infotron(s) supplied as clue(s), and results in brewing the
information required to satisfy the need of the information discoverer by utilizing the
documents available at its disposal (as information space). The various components of
the system and their interaction follows the penta-tier architectural model and therefore
can be considered fault-tolerant. IDS is generic in nature and therefore the characteristics
and the specifications were drawn up accordingly. Many subsystems interacted with
multiple infotron dictionaries that were maintained in the system.

In order to demonstrate the working of the IDS and to discover the information
without modification of a typical Library Information System (LIS), an Information
Discovery in Library Information System (IDLIS) application was developed. IDLIS is
essentially a wrapper for the LIS, which maintains all the databases of the library. The
purpose was to demonstrate that the functionality of a legacy system could be enhanced
with the augmentation of IDS leading to information discovery service. IDLIS
demonstrates IDS in action. IDLIS proves that any legacy system could be augmented
with IDS effectively to provide the additional functionality of information discovery
service.

Possible applications of IDS and scope for further research in the field are
covered.