# SPARK: A SYSTEM FOR PARALLEL REPRESENTATION OF KNOWLEDGE

Gerald A. Wilson
Computer Science Laboratory
Naval Research Laboratory
Washington, D.C. 20375

In the System for Parallel Representation of Knowledge (SPARK) the ingredient of concern is not the high-level, human-like, modeling of knowledge, but the compact, efficient, and effective internal representation and use of the knowledge. SPARK employs a knowledge base representation technique which has been shown to be as much as seven times more efficient for information retrieval than some other relational representations. At the same time, this technique, called the Parallel Representation (PAR) Technique, can also compact the knowledge base by a factor of two or more. What distinguishes SPARK from data management systems is that this efficient and effective retrieval mechanism also provides a powerful deductive inference capability.

Two types of parallelism are employed in SPARK, one achelved by data structures anc the other by parallel processing. Both are made possible by the distinction made between the "structure" and the "content" of data. In human problem solving the structure is the general concept while the content is the sets of items which, when combined with the structure, make one or more instances of that concept. Thus "transporting A from X to Y" is a concept while "carrying the block from the floor to the table" Is an instance of that concept with the content: carrying; block; floor; and table. For the basic constructs (the individual facts and inference rules of the knowledge base) PAR employs templates to represent the structure and sets to specify the content. For example, the collection of facts about objects supported by the table would be given by:

```
((R,X,Y){[supports]/R,[table]/X,
        [block,cone,lump,hammer] /Y»
```

where (R,X,Y) is the template specifying the concept of a binary relation with two independent arguments. The sets associated with R, X, and Y can be used to form specific instances by appropriate substitutions, in this case simply ordered cross products of the sets. To index the knowledge base PAR employs meta-templates and super-sets in a corresponding manner. The indexing structure proovides a compact form which facilitates efficient search and retrieval. Thus the representation Is parallel because any single symbol appearing in a PAR structure can represent an unbounded number of instances of that symbol in the knowledge base.

The second type of parallelism is multi-processing made possible by the meta-templates and super-sets of the index structure. The meta-templates are cannonlcal B-trees which partition the knowledge base into disjoint collections of data. When a query pattern matches n meta-tempates, n independent processes may be created to complete the retrieval match, thus performing many retrievals in parallel.

SPARK, with the PAR Technique, is not posed as a panacea for all knowledge base management problems. Several constraints were assumed in the development:

(1) Very large knowledge bases (more than 10**12 bits) are to be commonly employed.

(2) There is a significant degree of interrelationship among the elements of the knowledge base. If the knowledge base is viewed as a collection of n-tuples, then any distinct argument of a tuple has a high probability of appearing in multiple tuples.

(3) Search and retrieval are the preponderance of knowledge base operations.

(A) Search and retrieval may be equally likely for any combination of arguments, i.e. a query n-tuple may have instantiated any combination of argument positions, the remaining positions being left free.

(5) The representation must allow semantic (domain specific) constraints to be used In the search and retrieval process.

(6) Sets should be treated as sets.

(7) The representation should facilitate the use of inference.

These constraints appear to be quite general and representative of a large variety of realistic knowledge bases.

The Parallel Representation Technique employed in the SPARK system is posed as an approach to intelligent knowledge base management (i.e. management employing inference) for very large knowledge bases. Preliminary results from a simplified model and analysis of the technique indicate the potential for significant storage and search processing savings over some other relational representations. It is significant to note that the space savings due to the knowledge base compression ability of PAR do not cause an increase in the effort required to search the knowledge base on the average. The search mechanism can accomplish its task more efficiently in fact. This is due primarily to the elimination of any conflict between the manner in which the information is stored and the manner in which it is utilized by the search mechanism. Because the PAR Technique is intended as an internal representation of information it can be adapted to many different high level external representations.

The implementation and testing of SPARK is continuing at the Naval Research Laboratory. Once the system is fully operational experimentation will be made with large practical knowledge bases to further determine the strenghts and weaknesses of SPARK and the PAR Technique.