Scaling the Kowari Metastore

David Wood
MIND Laboratory, University of Maryland, College Park, MD 20742
dwood@mindswap.org

Abstract. The Kowari Metastore is an Open Source RDF database, built with the goal of providing large-scale storage of Resource Description Framework (RDF) information and a means to analyze that information in near real time. Kowari is currently the most scalable RDF database available. This paper presents a survey of options to further scale the Kowari Metastore. Options to further increase the upper bound of RDF statements which may be stored, decrease write times, distribute computational and storage resource requirements and improve analysis capabilities on large data sets are presented.

1 Introduction

The Kowari Metastore [12] provides a storage and analysis platform for RDF, RDFS and OWL information.

Several systems currently exist for the storage of RDF data. These databases – often referred to as RDF stores – exist as both Open Source projects and commercial product offerings. Since Guha’s first RDF store, rdfDB [3], several other Open Source stores have been developed. These include: Sesame [11], developed by Administrator Nederland b.v. as part of the European IST project On-To-Knowledge; 3store [10], developed by the University of Southampton; and Redland [2], developed by the University of Bristol. The Jena Semantic Web Framework [5], developed by Hewlett Packard’s Bristol Laboratory also provides the capability for persistent storage (thus making it an RDF store) by utilizing a persistent backend.

Kowari is currently the most scalable of the RDF stores. The RDF stores listed above allow for the storage of between hundreds of thousands and tens of millions of RDF statements, whereas Kowari is currently able to store hundreds of millions of RDF statements and query them rapidly. Figure 2 shows load times for 250 million statements on commodity hardware, although 500 million statements is currently considered a practical limit.

Kowari was developed by Tucana Technologies, Inc from 2001-2004. A commercial product, the Tucana Knowledge Server (TKS), is based upon Kowari and has recently been purchased by Northrop Grumman Corporation. Active development of Kowari continues as an independent Open Source project. [7]

Figure 1 shows Kowari's high-level architecture. The topmost layer provides access APIs, below which is a query engine, backed by a storage layer. The upper API
layer supports many of the common Semantic Web and industry standard access APIs. The Query Engine is separated from the access API layer by a transport layer, allowing the APIs to be used across a network connection. The Session and Resolver APIs are responsible for taking access requests and directing them to the storage layer.

**Fig. 1.** Simplified Kowari architecture

**Fig. 2.** Loading times for RDF statements in N3 format using manual commits on 64-bit and 32-bit machine architectures, noting the performance gains when mapped I/O is used.