Long-Term Research Plan (Ichiro Fujinaga)

Distributed Digital Music Library

My research plan for the next four to six years will concentrate on developing frameworks and tools for creating distributed digital music libraries.

It is becoming abundantly clear that the optimal path for creating useful sources of information is to distribute the task of digitizing the wealth of historical and cultural heritage material that exist in analogue formats, which may include books, manuscripts, music scores, maps, photographs, analogue tapes, and vinyl records.

In order to achieve this goal, libraries, museums, and archives throughout the world, big or small, need workflow management tools to digitize their collections and make them available efficiently.

In the past four years, I have been working with a group at the Johns Hopkins University for creating digital workflow management tools for digitizing their sheet music collection with the aid of various grants including NSF’s Digital Library Initiative Grant ($530,000 US) for which I was a Co-PI. I’m also the principal architect of a software, called Gamera, which originated as an optical music recognition system, but recently has been extended to create applications to recognize structured documents in general. This is one of the major components of the digital workflow management tools and I, as a Co-PI of the new $1,500,000US NSF Information Technology Research grant (2002–6), will continue to develop the software.

I am also waiting (to be announced 24 September 2002) for the result of a $250,000US grant from the Institute of Museum and Library Services to develop ways to digitize the audio tapes of Peabody Conservatory of Music’s concerts from the past forty years.

What is not directly investigated in both of the research above is how to access the data once they are stored. This problem arises even for new digitally-born materials.

Since we cannot expect thousands of libraries world wide to agree on a same database system or a query system, I will be investigating the possibility of applying the emerging technology called Web Services, which is designed to exchange information between different systems. The technology includes: Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL), and Universal Description, Discovery, and Integration (UDDI). Based on the XML format, the UDDI is used to register each institution’s services; WSDL is used to describe the type of service, access protocol, and its location; and SOAP is used as the protocol to exchange information.

The attractiveness of this technology for application in distributed digital libraries is that it assumes that each system (library) will be different. Web Services provide users with what, where, and how to access information from disparate systems.