Minutes

GGF9, October 05-08, 2003, Chicago, IL, USA
GridIR Working Group
October 08, 2003
GGF9, Chicago

Presenters:
Kevin Gamiel (MCNC-RDI/CNIDR)
Gregory Newby (Arctic Region Supercomputing Center)
Nassib Nassar (Etymon)

Minutes:
Sousan Karimi (MCNC-RDI/CNIDR)

Greg started the GIR-WG session by introducing himself, Kevin Gamiel,
Nassib Nassar and Sousan Karimi. He welcomed Nassib as the new
co-chair of the GIR working group. He then talked about the GGF
Intellectual Property policy, and mentioned that it can be found on
the GGF web site. He asked participants' permission to record the
session to help capture the meeting's minutes. There was no
objection. Greg mentioned the GIR mailing list, and asked those who
have not subscribed to it to do so. He informed the audience about
the teleconferences that have been hosted by MCNC-RDI. He said
teleconferences are always announced on the mailing list; and people
can send email to Sousan to obtain the phone number and the pin number
to join the call.

Greg talked about the meeting's agenda, and continued with the status
of the documents. He said the architecture document has changed some,
but not much. The requirements document was submitted; we had
positive feedback and will soon find out what the next step is. We
have not started the specification document yet but it is the next
document that we are going to work on. For that we need experience
gained from a reference implementation. Our reference implementation
is within the Globus Toolkit and OGSA framework; so we need to take a
look at all of the interaction that goes on within that framework. At
this time we have one reference implementation; however, we need to
have at least another one. We also have to look at what we need
conceptually to come up with an itemization of the specification.

Greg continued: Most likely by GGF10 we will have the architecture
document ready for last call. Nassib and I are also looking to start
on another reference implementation. GGF requires at least two
reference implementations and of course three or four would be better.

At this point Greg asked if anyone is interested in working on another
reference implementation, and he said if you are going to use Globus
we'd like to share our experience, especially Kevin's and the rest of
the CNIDR team, to help you with your efforts.

Then he went on to say that we are finding out (and you will hear
during the day) that GridIR is more of an application than
infrastructure, although it has some infrastructure components and
also some domain specific applications such as genomic
information. You will find out that there are few working groups that
are as far along as we are especially in terms of implementation based

on Globus3.0. This means that we may come up with some issues and questions for Globus developers and other working groups that are specific to our needs. We need to work as much as possible with other working groups and find the commonalities among our groups.

At this time Greg asked if there is anyone who works with any IR components that can be gridified? There were two positive responses, and he asked them to send email Kevin to continue the discussion. Then he asked the participants if any of them are involved in any other working groups that have some relationship to what we are doing, and added that it would be very beneficial if they share that information with the GIR-WG.

Greg stated that we need the reference implementation to be functioning software, and we need to address some of the challenges of GridIR. One of the hardest challenges, from the human information seeking point of view, is the query processing aspect. GridIR is very similar to WAIS, but in order to make it better than that we need to combine and merge the information that is coming back from many different sources in a meaningful and useful manner. Also we need to do relevance ranking correctly, and these are very exciting and open research areas. We also need other IR components, like the collection manager and indexers.

At this time he talked about the collection manager and indexer components of GridIR. (For more information refer to the requirements and the architecture documents available on www.gridir.org.)

Then he continued: The security issue is also very important, and even though there is the OGSA security, we need to solve more security related issues that are related to documents or in general to IR systems. There is also the matter of performance and accuracy. There are performance matrices in IR; however, there will be some issues related to the grid since the information is coming from multiple sources.

At this time Greg asked if anyone had any questions about IR or GridIR. There were no questions.

Lavanya: Have you talked to any one in DAIS working group for query processor or data transfer?

Greg: No, but we are following and tracking their work. Some working groups may have come up with for example a query processor, but most likely it will be very application specific and we cannot necessarily use that.

Kevin: We are going to use whatever comes out of any other group related to GridIR that we can take advantage of. For example the collection manager may be a candidate for that, and if we find out there is another group that has that, then for sure we will look into it and use it if it fits our requirements.

Lavanya: We also need to have joint group discussions that can be very useful.

An attendee: It looks like there are some common functionalities

between GIR and other working groups. Focus groups in a workshop would be good.

Kevin: We can have mini workshops in future GGFs.

An attendee: Next GGF has full day workshops.

Nassib: It will be great if some of you attend different related working groups and report back what you think is related to GIR.

Greg mentioned few of the working groups that attending them are beneficial to GIR-WG, and invited Nassib to start his presentation.

At this time Nassib started his presentation which was about SRW and Z39.50; you can find the presentation at: http://gridir.org/papers/gir-ggf9-zing.htm.

During the presentation Greg asked if anyone is familiar with Z39.50 and some answered, yes. He also said we are not saying that Z39.50 is the answer but we can find some of the answers in that. Greg also talked about the importance of some the functionalities such as explain in information retrieval because of the amount of data available in the information age.

Nassib mentioned that he has attended a meeting with the SRW group to find out if GIR can benefit from some of their work. He also talked about the significant effect of moving from static web services to dynamic grid services. Then he mentioned Matthew Dovey's suggestions and contributions to the architecture document (you can find details in the mailing list archive) and his participation in the SRW group. Greg mentioned that since SRW is an ongoing effort we should keep an eye on them, and Kevin stated that SRW is kind of oversimplified Z39.50 and it seems that now they have realized that.

An attendee: How much of these issues are relevant to IR vs. DAIS?

Nassib: For sure it is very relevant to IR.

Kevin: There are a lot of issues that can be discussed from the perspective of data access vs. IR. Z39.50 made most issues very complicated.

Nassib: It is harder to look at these in abstraction and it will be easier to see the issues when we do implementation.

An attendee: We have to find the common functionality among other groups and try to leverage each other's work.

At this time everyone showed agreement on finding out more about other groups and trying to work with them; there were other comments around the same issues. Greg pointed out that databases and IR systems can also work together and he said for example one of the IR systems that he works with uses database at the back end and IR for ranking and merging.

Kevin stated that at this time we are looking at SRW; however, we can work with different systems in parallel. He continued that we should

identify what needs to be done next and be active on the mailing list.

Next there was a discussion about natural language, and Nassib suggested that maybe we should add Question and Answer issues to the architecture document. Sousan asked if the GridIR query processor can
be used as an entry point for Q&A and Nassib stated that is how he sees it. Kevin said it can be done both ways, and he thinks the best solution is to have a specialized query processor that can be used as a plug-in to handle Q&A.

Kevin continued that we should identify those working groups whose work we need to follow and make sure someone from GIR-WG will work with them.

At this time the meeting came to the end and the attendees were encouraged to attend the next session that covered the demo of a GridIR reference implementation.