The Knowledge Grid Environment
Hai Zhuge

CS690 – Week 2 Slides

Jordan Osecki

Summarizing Key Ideas

- Describes a hypothetical future “Knowledge Grid Environment”, which is an autonomous human and machine combined environment which will evolve with the natural processes of science, technology, culture, and society
- Will consist of autonomous individuals, self-organized semantic communities, an adaptive-networking mechanism, and an evolving semantic networking mechanism
- Memex extension (ME) will offer a general model for autonomous individuals in this environment. These ME’s are configurable, adaptive, and context-aware digital organisms that will model network resources and host software and devices, organizing themselves to perform tasks based on socioeconomic principles
- ME’s will interact, inherit from each other, and work together to produce solutions, with little to no human guidance. These digital organisms will improve data flow and increase the productivity on the network

Memex Framework

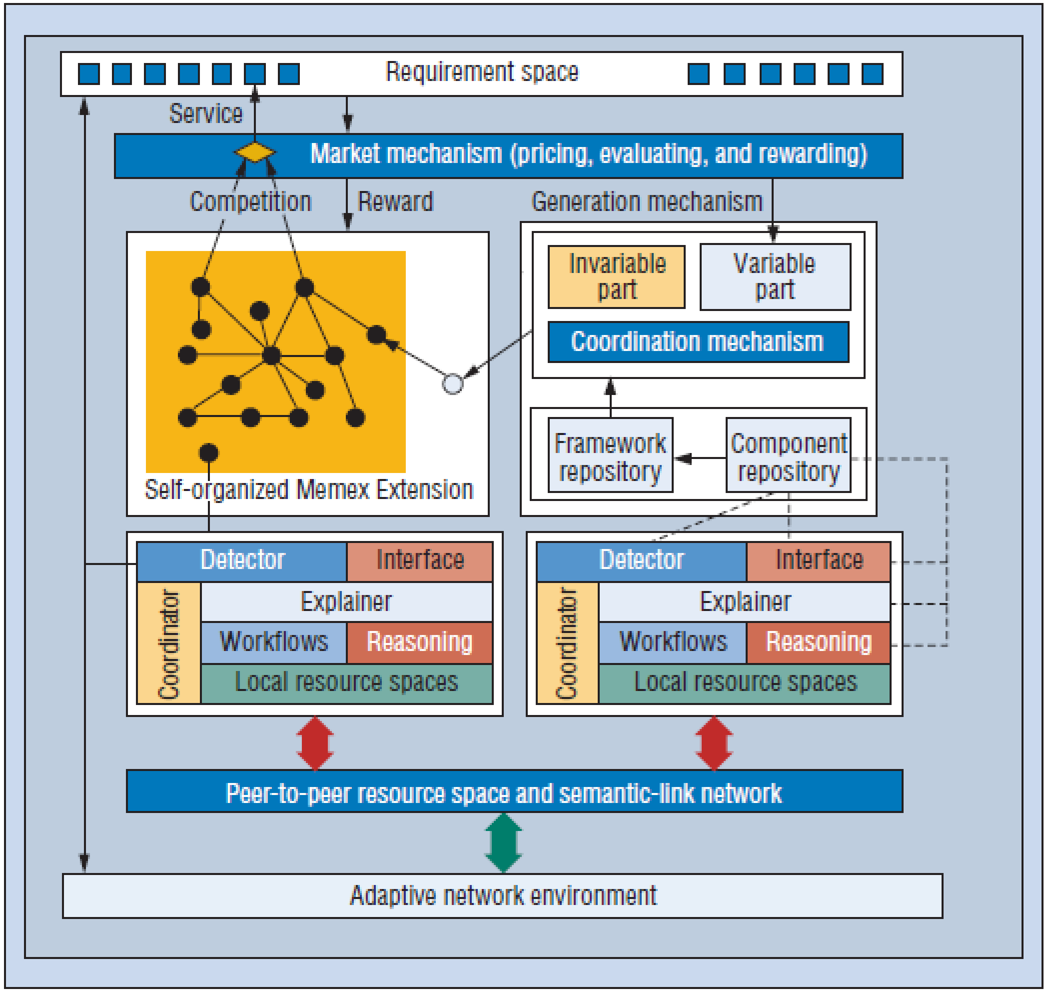

Figure 3. The memex extension (ME) world architecture.

Paintings Example

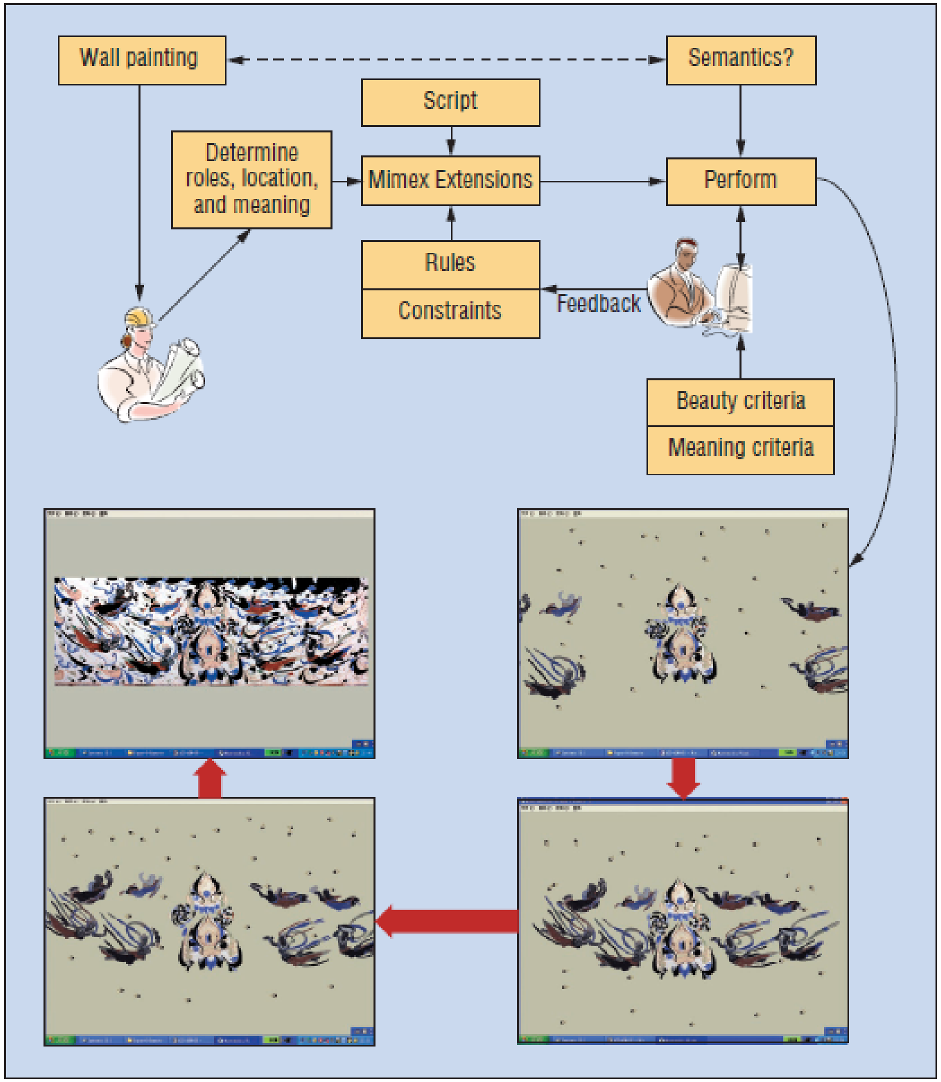

Figure 5. Animation process of Dunhuang wall paintings.

Workflow Planning on a Grid Thompson, C.W.; Wing Ning Li; Zhichun Xiao

CS690 – Week 2 Slides

Jordan Osecki

Summarizing Key Ideas

- Using workflows on grids in order to handle thousands to millions of files and terabytes of data
- Creating a prototype workflow software which automates the specification of workflows to avoid human error and ensure only legal workflows are ever used

Workflow in Grid

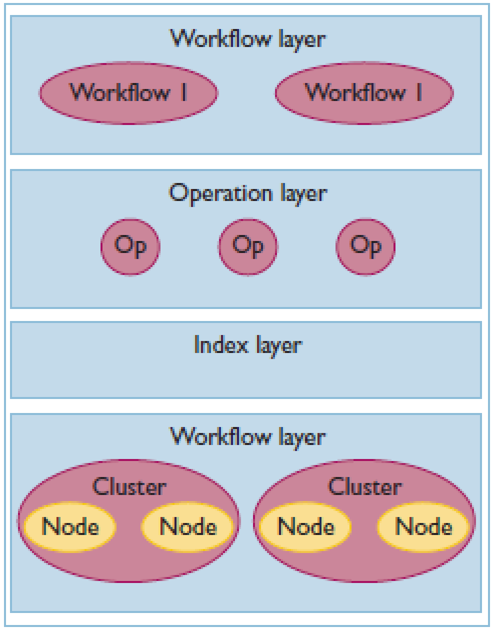

Figure 2. Workflow-management system. A workflow depends on operations. Many are custom and some are common relational operations that depend on an indexing layer for efficient implementation. Workflow operations and indices are implemented on top of a grid abstract machine consisting of nodes partitioned into clusters.

Prototype Example

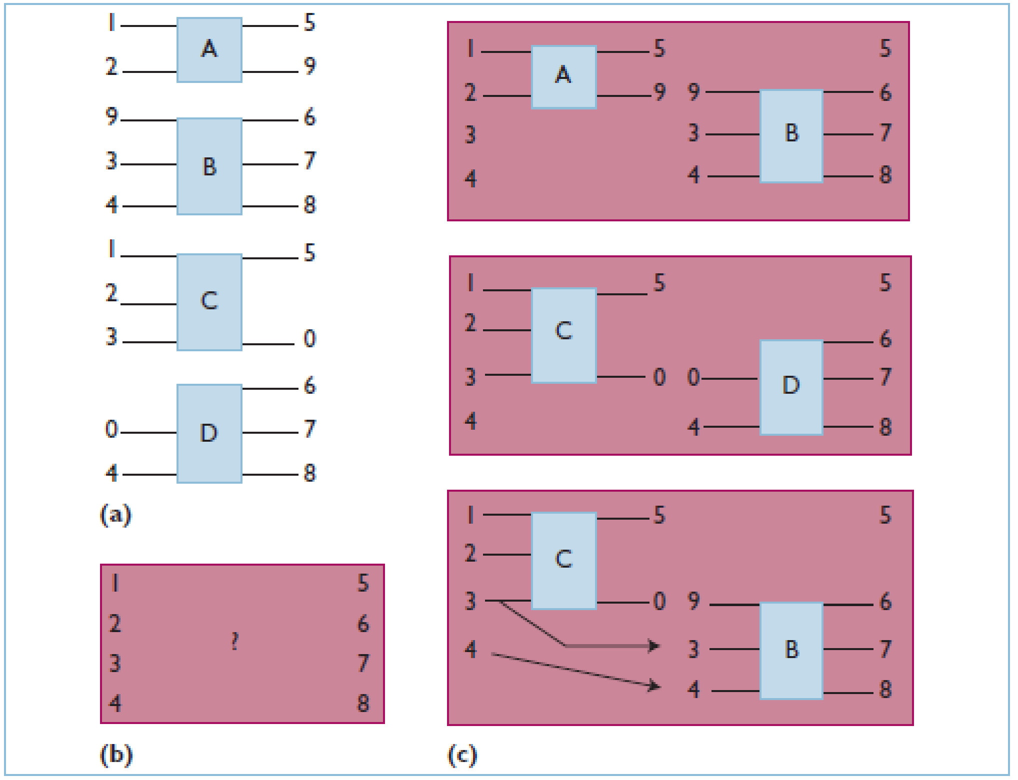

Figure 3. A prototype workflow planning system. This diagram shows (a) workflow operators A – D, (b) a workflow problem, and (c) how the operators can connect to solve the problem. For each operator, assume inputs on the left and outputs on the right, that both 0 and 9 can be matched to 9, and, for the rest, that a field can be matched only to itself.

Limitations

- Does not handle workflow operators whose parameters may change the operator’s specification
- Does not handle operators that map a signature of input fields to the same signature of output fields (sort)
- Does not handle algebraic equivalences for relational algebra operators (commutative)
- Does not handle cost-based optimization

- To solve, authors recommend trying to integrate with a relational system, but current ones aren’t open and flexible enough to accommodate highly parallel grid workflow operations dominated by domain-specific operators

E-Science: the grid and the Semantic Web
De Roure, D.; Hendler, J.A.

CS690 – Week 2 Slides

Jordan Osecki

Summarizing Key Ideas

- The authors propose to integrate Grids and the Semantic Web into a “Semantic Grid”
- Scientists and anyone involved in e-science would be the immediate beneficiaries of such a prototype
- Scientists need a combination of both large computing and data power, but also the ability to organize and distribute it – this solution provides both

Example - myGrid

- The myGrid team produces and uses a suite of tools designed to "help e-Scientists get on with science and get on with scientists"
- The tools support the creation of e-laboratories and have been used in domains as diverse as systems biology, social science, music, astronomy, multimedia, and chemistry. The tools have been adopted by a large number of projects and institutions

- The team has developed tools and infrastructure to allow:
  - the design, editing and execution of workflows in Taverna
  - the sharing of workflows and related data by myExperiment
  - the cataloguing and annotation of services in BioCatalogue and Feta
  - the creation of user-friendly rich clients such as UTOPIA
- These tools help to form the basis for the team's work on e-Labs.

- Source: http://www.mygrid.org.uk/

Challenges

- Realizing the network effect – creating an ability to accumulate large amounts of descriptive information about e-science artifacts and resources
- Moving beyond centralized stores – many successes are still in-house, which do not promote scientists having access to each other's work
- Automated assembly – automations must be built into the system so that the semantic processes will not hinder the speed of the grid
- Collaboration tools – tools need to be developed which will help the scientists collaborate on the semantic grid
- Grid meets physical world – there must be large emphasis placed on how humans and the grid will interact (interface)