CS891: Web Archiving Seminar Dissertationpalooza Presentation

Hussam Hallak

Unsupervised Small-World (USW)

A Framework for Web Object Self-Preservation

Charles L. Cartledge

Ph.D. in Computer Science

Old Dominion University

Norfolk, VA

August 2014

What is the question?

How to preserve digital data for the long term?

Why is it important?

- Our personal and cultural heritages are moving from analog to digital.
- Photo albums to image files in folders
- Cassette to CD, DVD, ...etc
- VCR to CD, DVD, ...etc

What have others done to solve the problem?

- Large archive/institutional centric approaches.
- Institutions can and will fail.

What will Dr. Cartledge do to solve the problem?

- Permeate the digital objects with the desire to preserve themselves and provide an environment where they can thrive
- USW is where graph theory, emergent behavior, and preservation meet.

How to prove that the problem is solved?

- The autonomic migration of digital/web objects from host to host and from one form to another.

What is the conclusion?

- Digital data preserves itself and outlives the people and organizations that created them.

Keeping something safe in the bad old days!

- Constant care & attention
- Don’t touch it
- Keep from the sun
- Keep from moisture
- Keep from other elements
- Keep from insects
- Keep from animals
- Keep from kids

Aleppo, gateway to castle

http://www.kacmac.com/cities/images/Aleppo/OldCaravaCitadel.jpg

Keeping something safe in the digital era!

- Refreshing: Preserving bits sequence.
- Migration: Preserving the content at a semantic level, but not necessarily the specific sequence of bits
- Emulation: Providing an environment where the original sequence of bits can be used

https://studentloanhero.com/featured/best-budgeting-apps-new-grads/

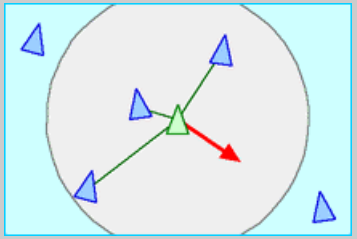
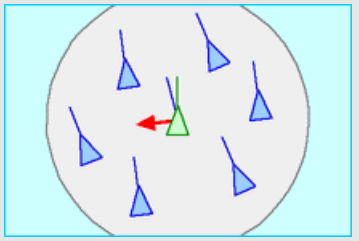
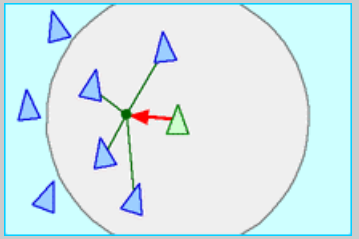
Systems others have created

- SAV: An architectural proposal
- aDORe: An operational system focusing on complex digital objects
- Fedora: A middle-ware systems ready to be the core repository technology in a local deployment with rich semantic object modeling for digital objects
- Dspace: A set of tools for the management of digital repositories, sponsored by MIT and HP Laboratories
- Lots of copies keeps stuff safe (LOCKSS), sponsored by the Stanford University Libraries
- Storage Resource Broker (SRB) (and its follow-on Integrated Rule-Oriented Data System, iRODS): An integrated rule engine with distributed storage and micro-services

What is emergent behavior?

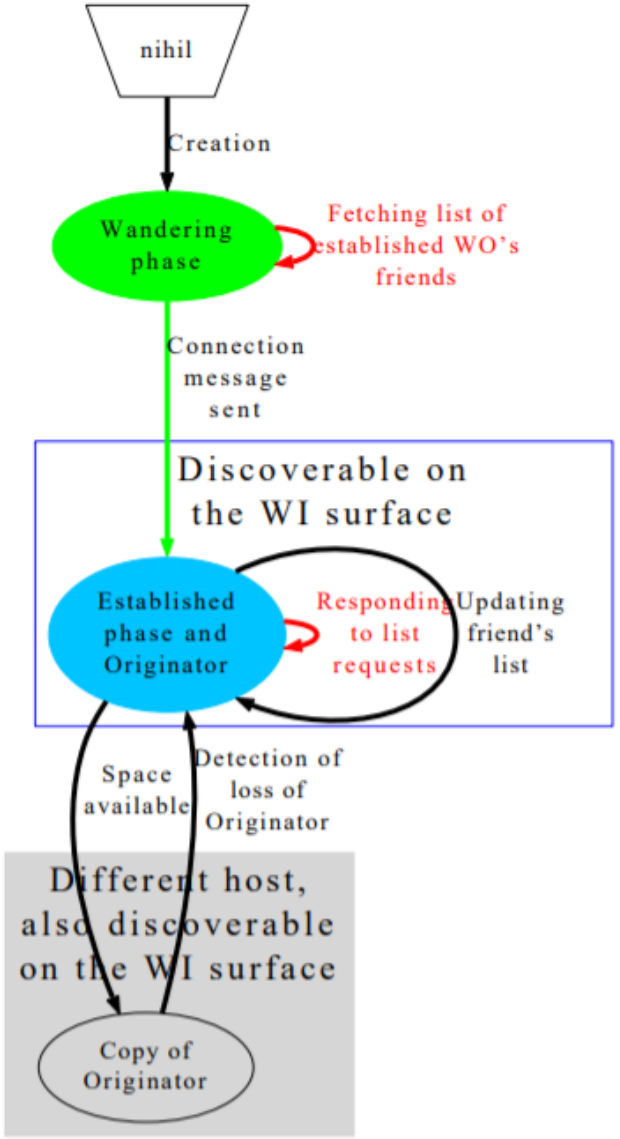
- Definition:
  Sometimes a system with many simple components will exhibit a behavior of the whole that seems more organized than the behavior of the individual parts. . . . Such phenomena are called emergent behaviors of the system.
- Examples:
  Schools of fish swim in harmony
  No collisions
  No enforcement of organization

Autonomous Group Control Philosophy

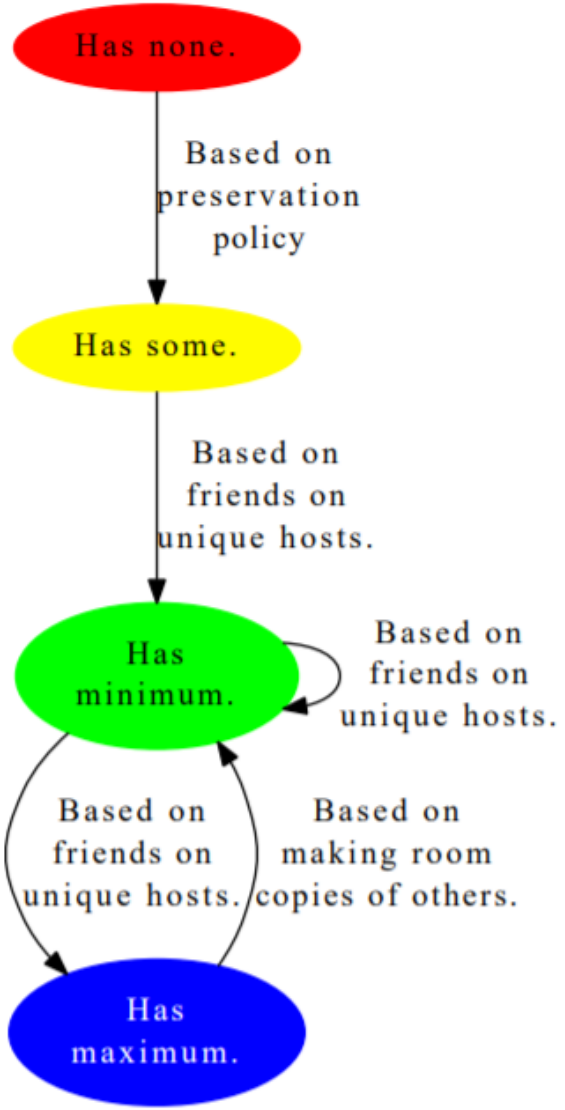
Reynolds' Concepts	Concept definition	Demo	As interpreted within the USW
Collision Avoidance	Avoid collisions with nearby flock-mates.		Each WO will have a unique name (a URI).
Velocity Matching	Attempt to match velocity with nearby flock-mates.		Resources use: as many resources as those in your "flock."
Flock Centering	Attempt to stay close to nearby flockmates.		Copies of a WO are near other WOs of the same "flock"; the preserved copy of a WO will only exist on a host known by a member of the flock.

http://www.red3d.com/cwr/boids/

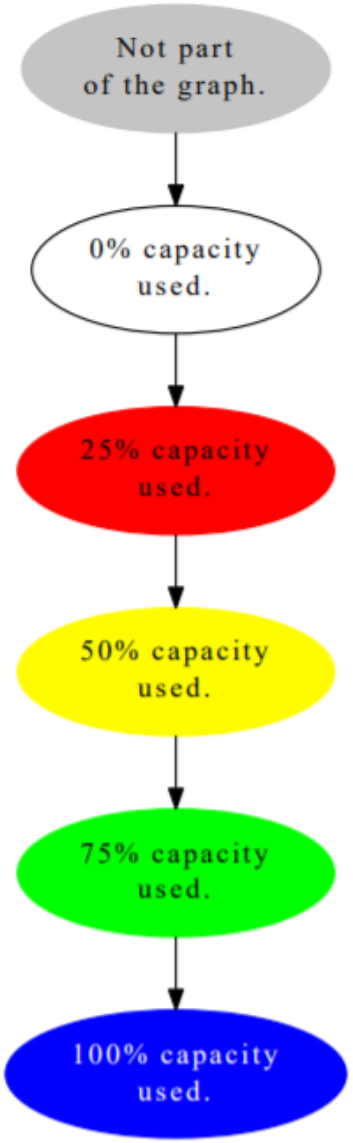
Life Cycle



WO's life cycle.



Preservation's life cycle.



Host's utilization.

Future work

- **Communication costs:** Not only to reduce the number of messages that are sent during the different phases of a WO's life, but also defining what is the "best" communications mechanism.
- **Graph theory:** Dealing with the actions that a USW graph must take when a previously disconnected graph elects new parent WOs and then becomes connected.
- **Information retrieval:** USW WOs will live a long time, but they have not fulfilled their purpose if their payload remains undiscovered and unretrieved.
- **Self modifying software:** USW WOs will live a long time, and it is reasonable to assume that the language that the original USW scripts were written in will become extinct. Therefore the scripts must be able to evolve to new languages and new capabilities.

References

- Chuck Cartledge (2014) “A Framework for Web Object Self-Preservation”