The Grand Challenges of Long Tail Learning and Technical Potentials

Roy Pea
Stanford University
CSCL09 Rhodes, Greece
11.06.09
Why the Long Tail Matters for Learning

- The potential of actualizing a cyberinfrastructure to support the vision of Dewey (1915) on *Interest and Effort* in learning; and Ivan Illich’s (1967) “learning webs” (communities of interest) in *Deschooling Society*.

- Successes in other sectors:
  - **Amazon**: BOOKS
  - **iTunes**: MUSIC BUYING
  - **NetFlix**: MOVIES
  - **Pandora**: MUSIC RADIO

- Personal - My lost ‘possible self’ as a Geologist

- Long Tail Learning as a *Grand Challenge Problem*: Prospect of personalized learning pathways for any topic at any level
"The genuine principle of interest is the principle of the recognized identity of the fact to be learned or the action proposed with the growing self; that it lies in the direction of the growing self...and is, therefore, imperiously demanded, if the agent is to be himself. Let this condition of identification once be secured, and we have neither to appeal to sheer strength of will, nor to occupy ourselves with making things interesting."
“[Consider] cases of indirect, transferred, or technically speaking, mediated interest. Things indifferent or even repulsive in themselves often become of interest because of assuming relationships and connections of which we were previously unaware. Many a student, of so-called practical make-up, has found mathematical theory, once repellent, lit up by a great attractiveness after studying some form of engineering in which this theory was a necessary tool.” (p. 22).

Dewey (1915) Interest and Effort
A passion for geology

• In 3rd grade – fell for rocks, minerals and gemstones.

• No teachers in my school with much knowledge.

• After endless pursuit of learning resources, and one neighbor’s rock collection and local knowledge, I exhausted my local libraries and accessible human help.

• Gave it up.
What’s different about learning resources than typical long-tail phenomena?

• Ontology: Not straightforwardly isolatable products in the manner of books, songs, movies

• Textbooks? No – finer level granularity needed.

• Learning objects? Little success so far.

• Websites? Complex objects with multiple pages and navigational structure.

• Need resources appropriate to developmental, educational and reading levels – not only by topic.
Is there a problem – or what is the problem?

- How useful to learners of all ages is Google? Yahoo? The Open Directory Project?
- Issue is not whether a learner can find relevant learning resources.
- Problem is finding the best-matched learning resources to your interests with the least effort.
- But - a learner’s interests may often be vague in details, and appropriate resources may be tacitly constrained by their level of reading competency or conceptual development.
- Which implies the need to track implicitly the resources they are using & text that they are writing, to infer learning interests and competencies.
Potential Resources

• Wikipedia entries as an ontology
  – 2,907,504 articles in English

• Open Directory Project (dmoz)
  – 4,616,309 sites - 83,367 editors - over 590,000 categories
  – ODP uses a hierarchical ontology scheme to organize sites. Listings on a similar topic are grouped into categories, which can then include smaller categories.

• Mirrored and expanded by Google with its PageRank algorithm
Wikipedia: A place to start?

- Millions of articles on named topics
- Could users select a ‘Learning station’ like they select a ‘Pandora station’ – starting from a topic in Wikipedia, and having new learning topic entries queued up like songs on the personalized Internet radio station Pandora?
- Need an ontology for learning topics like the musical genre components of Pandora.
- Learning stations for individuals would be refined with thumbs up/down as users experience the recommended learning resource.
Imagine: Multi-Voice Interest Profiling

• The limitation of recommendation engines is the assumption that any given individual only needs one profile.
• But individuals have multiple roles, persona, or ‘selves’ that have different informational and affiliative needs.
• This service would allow a learner to specify different nameable profiles or “voices” that they can distinctly use to generate preferences and to seek recommendations affiliated with each preference set.
Problems with a single identity

• I may have one voice as a learning researcher in what books I like to read, another as a vacationer, and a third when I seek books for my daughter Elle.

• Yet with today’s recommendation services, my ‘preferences’ across these multiple roles are aggregated, and I find myself receiving recommendations for books like those I bought for Elle.

• This skewing of recommendation services created by the ‘noise’ of confounding different multiple ‘voices’ could be resolved by allowing anyone to specify a named identity for any distinctive interest profile they wish to define.

• …which creates an intriguing technical possibility…
…Shareable Interest Profiles

• Related to handing a student a book, or emailing a URL of interest.

• One cannot always ask a colleague or friend for a learning resource – so how to virtually ask for advice when one cannot make a personal, live contact with that person? (“Buddy, can you spare…a link?”)

• This service would allow a person to share any of their multiple-voice profiles with others who they allow to ‘read’ a subscription to their recommendation profiles.

• A ‘borrower’ of one's profile would receive recommendations from sites that have preference profiles & recommendation engines to receive the recommendations appropriate to one’s virtual buddy.
Complications

• Purpose of learning
  – Learning that (facts, conceptual networks)
  – Learning how (skills)
    • Thread a needle; make sushi; putting; sword-making
  – Learning a disposition
    • Meditative attitude, mindfulness; for social organizing

• Multisensory modalities of learning (>text)
  – Oral and musical (rhythm; melody)
  – Visual discrimination (color; decorating)
  – Smell and taste (preparing wine and food combos)
  – Tactile and haptic (Knitting; knot-tying; billiards)
Long Tail Learning is Complicated

- 6 billion people, each with many many interests, for many purposes, with many sensory modalities, and using many different languages
- 5 billion or more webpages; the ‘deep web’ is far larger
- Long Tail Learning is so complicated, it engages two different Grand Challenge Problems for Cyberlearning

**Cyberlearning:**

- “Learning that is mediated by networked computing and communications technologies”
- Learning in a networked world, where the forms of “steering” of learning can arise in a hybrid manner from a variety of personal, educational, or collective sources and designs
Grand Challenge Problems: History

- **A grand challenge** defines a commitment by a scientific community to work together towards a common goal - valuable and achievable within a predicted timescale.

- **Predecessor**: Hilbert’s 1900 address to International Congress of Mathematicians on 23 major mathematical problems to be studied for the next century.

- **“Grand Challenges”**: major problems of science and society whose solutions require 1000-fold or greater increases in the power and speed of supercomputers and their supporting networks, storage systems, software and virtual environments:
  - *U.S. High Performance Computing and Communications program* (HPCC, started 1991)
Grand Challenge Problems: Environment

• 2001: National Academy Press publishes *Grand Challenges in Environmental Sciences* to identify high-priority environmental science projects for the next generation: “areas of opportunity that—with a concerted investment—could yield significant new findings.”

• Nominations for environmental science “grand” challenges solicited from thousands of scientists worldwide

Eight grand challenges were defined; the first four were recommended for most immediate research investment and attention

- Biological Diversity and Ecosystem Functioning
- Hydrologic Forecasting
- Infectious Disease and the Environment
- Land-Use Dynamics
- Biogeochemical Cycles
- Climate Variability
- Institutions and Resource Use
- Reinventing the Use of Materials
Grand Challenge Problems: Cyberlearning

• *Much is different* about grand challenges in cyberlearning than in math, computing, environment:
  – Education is *normative*, with *diverse standards* globally.
  – Fragmented researchers and knowledge, rare “collaboratory” arrangements for collective intelligence and action.

• *Yet much is the same* for Cyberlearning:
  – We need *our own* GCPs at the intersection of developments in cyberinfrastructure, learning sciences, and education.
Cyberlearning Grand Challenge Problem Criteria

1) **Understandable, with Significance.**
   Clearly stated compelling case for contributing to long term benefits for science, industry and society.

2) **Challenging, and Timely.**
   Hard problems within conceivable reach in 15-20 years with concerted coordinated efforts.

3) **Clearly useful, in terms of Impact and Scale, if problem is solved.**
   Contributes to long term benefits for many people at large, and with international scope.

4) **Metrics: Testable and Incremental.**
   Can measure progress, incremental milestones.

---

Jim Gray:
Director, 
Microsoft Research Lab, 
San Francisco


Inspiration
Cyberlearning Grand Challenges (Pea, 2007)

- Developmentally-Relevant Learning Support
- Tacit Interest Analysis
- Systems for Learning Performance Certification and Reputation Management
- Lifelong Digital Learning Portfolios
- An informed digital earth
- Language-to-Language Translation Capabilities
- “Reflexive' cyberinfrastructure systems
- Simulating learning with cyberinfrastructure systems
Cyberlearning Grand Challenges: Candidates (1-2 of 8)

1. **Developmentally-Relevant Learning Support:** Provide real-time access for any learner—whatever their current developmental capabilities—to relevant cyberinfrastructure support that can guide them toward meeting any learning goal. For any desired performance competency, configure in real time a distributed learning system that will discover requisite learning resources and, if needed, peers or mentors with verifiable reputations to help one attain such competencies in a certifiable manner.

2. **Tacit Interest Analysis:** Diagnose any learner’s interests from digital capture and analysis of what they read, talk about, and attend to, so as to automatically create an *interest profile* that can be used for compiling engaging content and scenarios for their pursuit of enhanced skills and competencies using cyberinfrastructure. Ensure appropriate privacy safeguards.
Summary

• Long-tail learning needs driven by learner interests
• Ontology problems of long-tail learning significantly different than for music, movies, books
• Technical prospects sketched for:
  – Leveraging Wikipedia, Open Directory services to create user-refinable ‘Learning Stations’
  – Multi-Voice Interest Profiles
  – Shareable Interest Profiles for receiving recommendations aligned to the interests of others
• 2 Grand Challenge Problems for Long Tail Learning