



InfoGarden:

A Casual-Game Approach to Digital Archive Management

Carlos Maltzahn, Michael Mateas, Jim Whitehead
Computer Science Department
University of California Santa Cruz
{carlosm,michaelm,ejw}@cs.ucsc.edu

Problem

- **People overwhelmed by personal digital archives**
 - ▶ Exponential growth faster than expected [Gantz+ 2008]
 - ▶ Digital archives already dwarf physical collections
- **Digital archives require attention for survival**
 - ▶ Digital archives won't run out of space: easy storage
 - ▶ Digital archives won't survive inattention: easy data loss
 - ▶ Digital preservation requires continual effort
- **Attending to digital archives is tedious**
 - ▶ Current approaches require significant amount of attention
 - ▶ Users don't seem to have time for that

So what?

- **Digital Dark Age**
 - ▶ Massive loss of personal documents common
 - ▶ Future historians deprived of documents essential for historic understanding
- **Compounding effect of poor data management on cost of:**
 - ▶ Data safety: more needs to be backed up
 - ▶ Data security: more needs to be encrypted

Game Reward Hypothesis

- Idea: Recast archive maintenance as fun & entertaining game
- Examples of productive games:
 - ▶ Chao's PSDoom [CHI'01]
 - ▶ Von Ahn's ESP [CHI'04] and other "games with a purpose"
- 2007 Casual Games Association report: 7-15 hours of play a week
 - ▶ Plenty of time for archive maintenance!

A game approach significantly increases time users dedicate to archive maintenance.

InfoGarden

- Focus on common & tedious activity: **tagging**
- **Idea:** personal archive ≈ garden



Neglected archive

Well-maintained archive

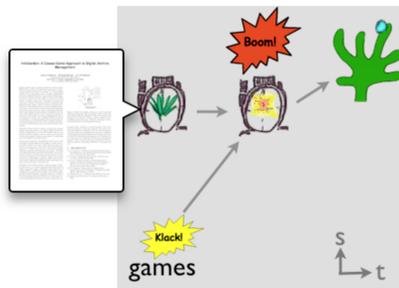
- **Goal:** weed and plant your way to a beautiful garden!

Semantics

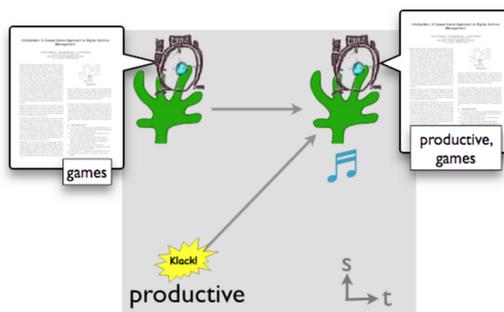
- **Weed:** document without tags
- **Plant:** taxonomy (tag hierarchy)
- **Fruit:** document / taxonomy association
- **Crosshairs:** reveals documents, taxonomies



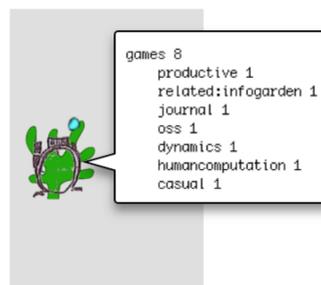
Shooting weeds with words



Shooting fruit with words



Revealing taxonomy

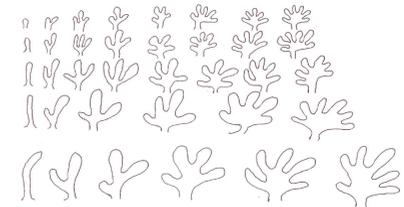


Scoring

- Shot weed = 1 + [new fruit] points
 - ▶ Plus value of tagged file!
- Deleted weed = 1 point
 - ▶ Minus value of file!
- Shot fruit = 1 point
 - ▶ Encourages focus on untagged documents

Plant Morphology

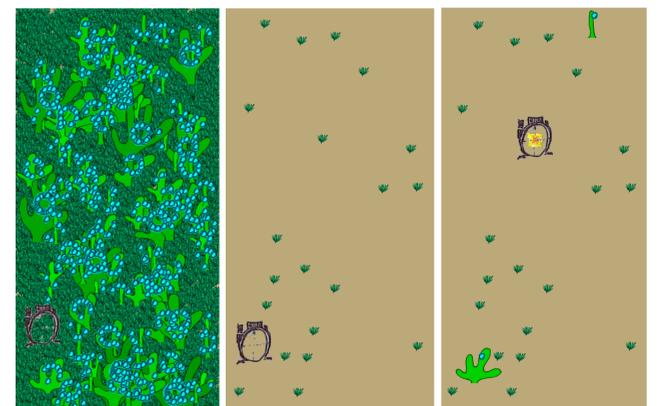
- **Plants** are tag taxonomies based on subset relationships
 - ▶ Subtag of Tag: $Docs_{Subtag} \subset Docs_{Tag}$
 - ▶ **Plant size:** $\log_2 |taxonomy|_{docs}$
 - ▶ **Plant branching:** $\log_2 |taxonomy|_{tags}$



Planting & Fertilizing

- **Planting:** use new tags without subset relationship
 - ▶ Example: $d1(t1), d2(t1, t2), d3(t1, t3)$
 - ▶ $t1: \{d1, d2, d3\}, t2: \{d2\}, t3: \{d3\}$
 - ▶ $d4(t4) \Rightarrow$ new plant!
- **Fertilizing:** shoot weeds with mixture of old and new tags
 - ▶ Old keywords *increase size* or *split plants*
 - ▶ Combination of old and new tags *increase branching* or *split plants*
 - ▶ Example:
 - ▶ $d1(t1, t4) \Rightarrow t4: \{d1, d4\} \Rightarrow$ d1 fruit on old and new plant!

Scalability



Showing Everything: Overwhelming!

Showing 20 weeds: Much better...

Plants show up: Based on fruit.

Summary & Future Work

- **Summary:**
 - ▶ Gardening game metaphor works for tagging documents
 - ▶ Preliminary experience supports game reward hypothesis
 - ▶ Simple scoring metric encourages orthogonal tagging
- **Status:**
 - ▶ Game works as described (Python, PyGame, Mac OSX)
 - ▶ Simple scoring: encourages orthogonal tagging
 - ▶ Uses standard Mac OSX Spotlight and QuickView features
 - ▶ Research platform for game mechanics
- **Future:**
 - ▶ **Get funding!**
 - ▶ Focus on game mechanics (e.g. threats and attacks)
 - ▶ Include plant morphology distribution into scoring
 - ▶ Apply to other areas of information management and analysis

