Experiment Explorer: Lightweight Provenance Search over Metadata

Presented by Del Davis

Written by:

- Delmar B. Davis (UWB)
- Hazeline U. Asuncion (UWB)
- Ghaleb Abdulla (LLNL)

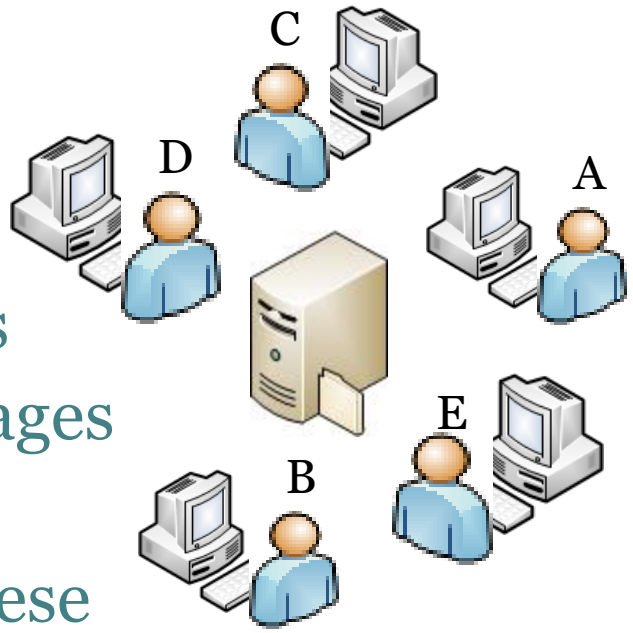
Lawrence Livermore National Laboratory

Motivation

Researchers need the ability to cross-reference data

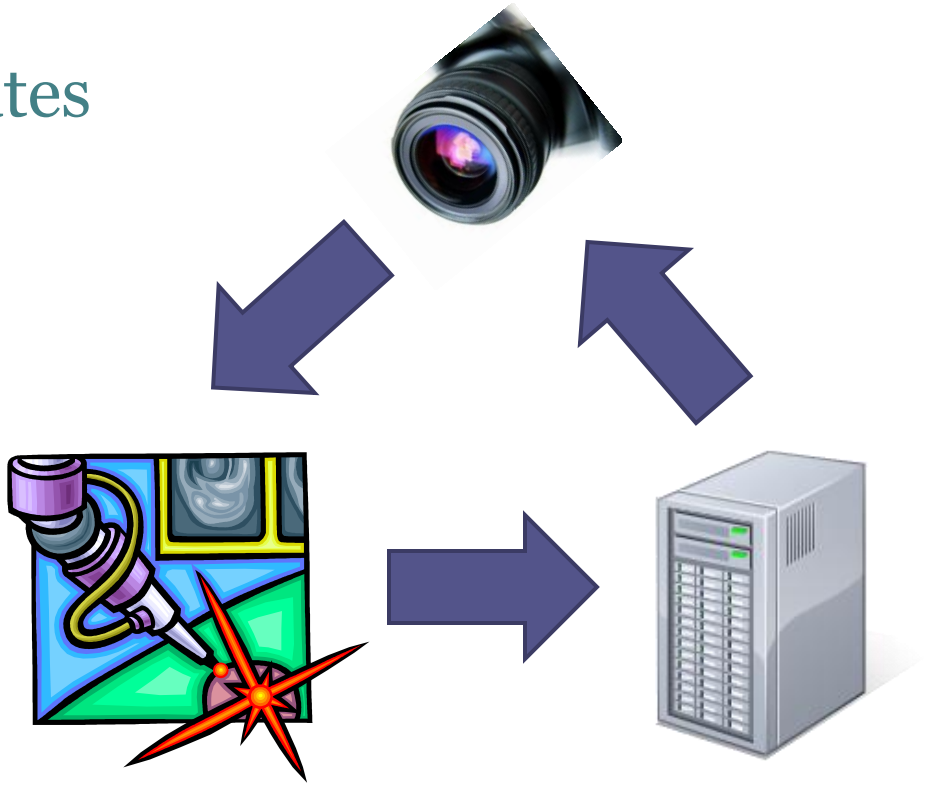
Scenario:

- A runs an experiment
- B finds an image from it
- C, D, and E complete experiments
- A extracts features from these images
- A and B analyze these features
- C uses reduced data from all of these artifacts in a publication.



Experiment Process at LLNL

- Create experiment design
- Build provenance templates
- Create workflows
- Run experiment
- Record output
- Analyze results



Overview

- Problem
  1. Many experiment artifacts are produced at LLNL
  2. Artifacts are located in various file systems
  3. Artifacts are in various formats
- Solution Requirements
  1. The solution must scale to the search scenario
  2. It must operate on a common view of the file systems
  3. Common fields are needed for queries

Search Techniques



- Desktop Search
  - Application recorded metadata
  - Full-text analysis
  - OS provenance

Search Techniques



- Desktop Search
  - Application recorded metadata
  - Full-text analysis
  - OS provenance
- Web Search
  - Implicitly linked documents
  - Cached searches

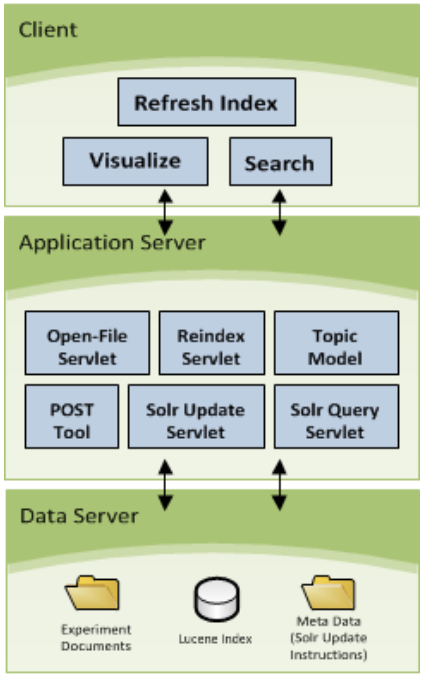
Search Techniques

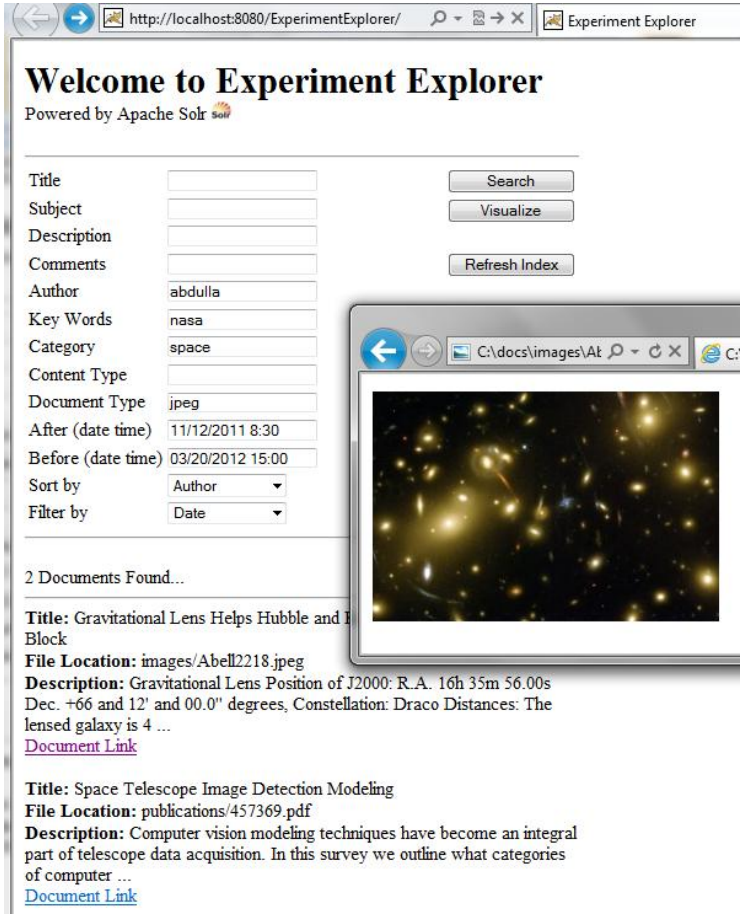


- Desktop Search
  - Application recorded metadata
  - Full-text analysis
  - OS provenance
- Web Search
  - Implicitly linked documents
  - Cached searches
- Database Search
  - Structured queries

Experiment Explorer Approach

- Metadata search





Experiment Explorer Approach

- Metadata search
- Topic modeling



New York Times Article (#39):

The attacks[45] of Sept.[37] 11 were aimed[20] at the symbolic[49] heart[50] of capitalism,[42] but that does not mean[50] the titans[6] of industry[15] feel[50] obligated[12] to join[26] the fight[50] on terrorism.[37] On numerous[49] occasions[50] since the attack,[45] businesses[15] have found[50] themselves facing[15] a choice[6] between[20] patriotic[3] action[50] or their bottom[26] line.[50] Politicians[20] have suggested[50] that investment[15] houses[4] would be caving[20] to terrorism[45] if they did not rebuild[9] in New York,[50] for example, and asked[50] domestic[37] wheat[6] growers[20] for understanding[48] as they purchased[35] Asian[20] wheat[6] for rapid[49] food[50] shipments[21] to Afghanistan.[12] They have encouraged[26] American[38] apparel[35] manufacturers[49] to continue[50] to buy[15] goods from Pakistan[45] even in the face[50] of rising[28] insurance[15] costs[15]

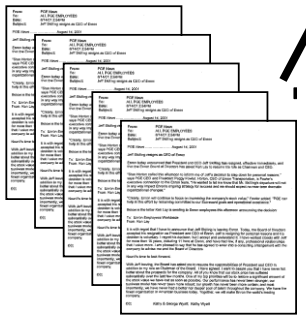
New York Times Article (#16):

But heading[29] into Sunday's[29] game,[29] they also see them as an inspiration.[49] The two teams[29] have gone in opposite[2] directions[1] since the Cowboys[4] got their first victory[29] of the season[29] over the Redskins,[29] 9-7 six weeks[3] ago.[15] That loss[13] was the fifth in a row[29] for the Redskins,[29] who were then billed[50] as a lost[4] cause. Now the Redskins[29] are talking[50] playoffs[29] after becoming the first team[29] in NFL[13] history[50] to win[29] five consecutive[29] games[35] after opening[29] the season[22] with five losses.[13] The Cowboys[4] won[4] one more game[29] after the victory[13] over the Redskins,[29] but are riding[50] a four-game[13] losing[13] streak[29] going[29] into the rematch[13] in Washington.[13] Cowboys[4] coach[29] Dave[13]

Topic Modeling

Topic Learning Algorithm

Text Data



Discovered Topics

[12]: war military force american afghanistan government united army
[13]: yard bowl team game season coach football quarterback
[15]: company enron companies million business stock market deal
[29]: game play season games team goal against player
[36]: taliban afghan afghanistan fighter forces bin qaida laden
[37]: american sept attack war world terrorist america united
[40]: bush president house bill republican administration congress senate
[45]: palestinian attack official laden bin terrorist pakistan group

New York Times Article (#39):

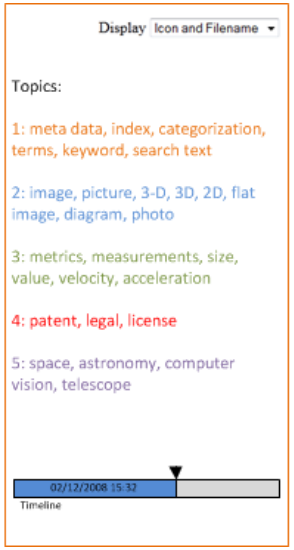
The attacks[45] of Sept.[37] 11 were aimed[20] at the symbolic[49] heart[50] of capitalism,[42] but that does not mean[50] the titans[6] of industry[15] feel[50] obligated[12] to join[26] the fight[50] on terrorism.[37] On numerous[49] occasions[50] since the attack,[45] businesses[15] have found[50] themselves facing[15] a choice[6] between[20] patriotic[3] action[50] or their bottom[26] line.[50] Politicians[20] have suggested[50] that investment[15] houses[4] would be caving[20] to terrorism[45] if they did not rebuild[9] in New York,[50] for example, and asked[50] domestic[37] wheat[6] growers[20] for understanding[48] as they purchased[35] Asian[20] wheat[6] for rapid[49] food[50] shipments[21] to Afghanistan.[12] They have encouraged[26] American[38] apparel[35] manufacturers[49] to continue[50] to buy[15] goods from Pakistan[45] even in the face[50] of rising[28] insurance[15] costs[15]

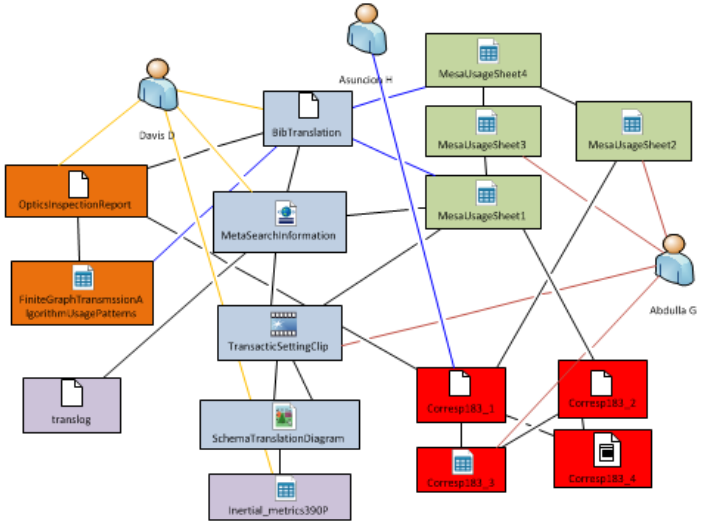
New York Times Article (#16):

But heading[29] into Sunday's[29] game,[29] they also see them as an inspiration.[49] The two teams[29] have gone in opposite[2] directions[1] since the Cowboys[4] got their first victory[29] of the season[29] over the Redskins,[29] 9-7 six weeks[3] ago.[15] That loss[13] was the fifth in a row[29] for the Redskins,[29] who were then billed[50] as a lost[4] cause. Now the Redskins[29] are talking[50] playoffs[29] after becoming the first team[29] in NFL[13] history[50] to win[29] five consecutive[29] games[35] after opening[29] the season[22] with five losses.[13] The Cowboys[4] won[4] one more game[29] after the victory[13] over the Redskins,[29] but are riding[50] a four-game[13] losing[13] streak[29] going[29] into the rematch[13] in Washington.[13] Cowboys[4] coach[29] Dave[13]

Experiment Explorer Approach

- Metadata search
- Topic modeling
- Visualization of results.





Acknowledgements

- This work is supported in part by the University of Washington Royalty Research Fund No. A65951.

Thank you for your time