Revealing the Copy and Paste Habits of End Users

*Kathryn T. Stolee, Sebastian Elbaum, and Gregg Rothermel

University of Nebraska-Lincoln

This work was supported in part by the EUSES Consortium through NSF-ITR 0324861 and 0325273, and CFDA#84.200A: Graduate Assistance in Areas of National Need (GAANN).
What We Do Not Know

(And why it matters.)

- How frequently is the clipboard used?
- How is the clipboard used across applications?
- In what contexts is data transferred via the clipboard?
Methodology

• Build monitor for clipboard activity
  – Works in Windows XP and Vista

• Distribute tool to end users
  – 15 users, ~50 hours per user

• Process the data
  – 2544 interactions (3% cuts, 45% copies, 52% pastes)
  – Average clipboard interactions per hour: 3.4
How is the Clipboard Used Across Applications?
Application Usage

98% of clipboard interactions \textit{per user} occurred within 6 applications.
## Types of Applications

<table>
<thead>
<tr>
<th>Type</th>
<th>Percent of Use</th>
<th>Percent as Source -- Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Processors</td>
<td>26%</td>
<td>36% -- 64%</td>
</tr>
<tr>
<td>Web Browsers</td>
<td>23%</td>
<td>58% -- 42%</td>
</tr>
<tr>
<td>Email Clients</td>
<td>19%</td>
<td>49% -- 51%</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>18%</td>
<td>51% -- 49%</td>
</tr>
</tbody>
</table>
Application Interactions

- 43% of pasted data had been copied in a different *application*
In What Contexts is Data Transferred via the Clipboard?
### Elementary Patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Graph</th>
<th>Overall Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td><img src="A-%3EB" alt="Diagram" /></td>
<td>65%</td>
</tr>
<tr>
<td>Within</td>
<td><img src="A" alt="Diagram" /></td>
<td>35%</td>
</tr>
</tbody>
</table>
Repeat Pattern

<table>
<thead>
<tr>
<th>Graph</th>
<th>Overall Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Graph Diagram" /></td>
<td>37%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Street Address</th>
<th>City</th>
<th>State</th>
<th>Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lincoln</td>
<td>Abraham</td>
<td>9 Alphabet Drive</td>
<td>Lenexa</td>
<td>KS</td>
<td>66215</td>
</tr>
<tr>
<td>Roosevelt</td>
<td>Franklin</td>
<td>123 Butterfly Drive</td>
<td>Penham</td>
<td>IN</td>
<td>66873</td>
</tr>
<tr>
<td>Washing</td>
<td>Lin</td>
<td>625 West 12</td>
<td></td>
<td></td>
<td>61502</td>
</tr>
</tbody>
</table>
Distribution Pattern

Graph

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
</table>

Overall Usage

32%
Composition Pattern

<table>
<thead>
<tr>
<th>Graph</th>
<th>Overall Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>18%</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

**Graph:**
- A
- B
- C
- D

**Overall Usage:** 18%

**Table:**

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Earnings</th>
<th>Diff from Last Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50K</td>
<td>4K</td>
</tr>
<tr>
<td>2</td>
<td>48K</td>
<td>2K</td>
</tr>
<tr>
<td>3</td>
<td>52K</td>
<td>2K</td>
</tr>
<tr>
<td>4</td>
<td>100K</td>
<td>2K</td>
</tr>
</tbody>
</table>
A Closer Look at the Complex Patterns
Distribution Pattern – 32%

Data Flow

Changes in Window Focus
Composition Pattern – 18%

Data Flow

Current Clipboard

B

C

D

X

Y

Z

A

Multiple Clipboard Buffers

B

C

D

X

Y

Z

A

Changes in Window Focus

Copy X

Copy Y

Copy Z

Paste X

Paste Y

Paste Z

Copy X

Copy Y

Copy Z

Paste X

Paste Y

Paste Z

Add x y and z
Repeat Pattern – 37%

Current Clipboard

Data Flow

A

B

Copy X

Paste X

Copy Y

Paste Y

Copy Z

Paste Z

Changes in Window Focus

Context Sensitivity: Term Extraction & Multiple Pastes

A

123 Road
City, ST

B

123 Road
City, ST

Copy XYZ

Paste X, Y, Z
Repeat Pattern – 37%

Current Clipboard

Data Flow

A → B

Iteration

A → B

Copy X
Copy Y
Copy Z

Paste X
Paste Y
Paste Z

Copy [X … Z]

Paste [X … Z]
Distribution Pattern, Revisited

### ESTIMATED EXPENSES

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATE VEHICLE</td>
<td>£ 5.00</td>
</tr>
<tr>
<td>PERSONAL VEHICLE</td>
<td>£ 515.00</td>
</tr>
<tr>
<td>AIRFARE</td>
<td>£ 370.00</td>
</tr>
<tr>
<td>LODGING</td>
<td>£ 436.00</td>
</tr>
<tr>
<td>MEALS</td>
<td>£ 200.00</td>
</tr>
<tr>
<td>REGISTRATION FEES</td>
<td>£ 96.00</td>
</tr>
<tr>
<td>TAXI &amp; FARES</td>
<td>£ 96.00</td>
</tr>
<tr>
<td>OTHER EXPENSES</td>
<td>£ 1,622.00</td>
</tr>
<tr>
<td><strong>TOTAL ESTIMATED</strong></td>
<td><strong>£ 1,622.00</strong></td>
</tr>
</tbody>
</table>

### Screenshot of Software Interface

- **Create Expenses**
  - **Planned Conditions**
  - **Details for Copying**
  - **Expected Savings**
  - **Net Value**
  - **Cone.**
  - **Currency**
  - **Priced at $**
  - **Exchange Rate**
  - **Balance**

- **Allocated Advances**
  - **Total Amount**
  - **Unallocated Amount**

- **Allocation of Balance**
  - **Due To E.I.T.:** £ 6.00
  - **Due To Traveler:** £ 200.00
Implications
## Table of Tool Support

<table>
<thead>
<tr>
<th>Feature</th>
<th>Citrine</th>
<th>Clip-Mate</th>
<th>MS Office</th>
<th>Quick Click</th>
<th>Restack &amp; Roll</th>
<th>Topes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Works with all applications</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Holds multiple items</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Extract context from data</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Facilitates multiple pastes at once</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Iterates through multiple items</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduces changed in window focus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helps manage multiple windows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>User-defined formatting support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Searches for type errors in dest.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Searches for context dependencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Represents data as objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keeps track of provenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>
Conclusion

• We studied how end users use the clipboard
  – End users copy and paste within a few applications
  – End user behavior is repetitive
  – A small set of patterns represents end user behavior

• We identified several support features
  – Context-sensitivity
  – Sequential iteration
  – User-defined formatting support
  – Search for type errors in destination
  – ...

21
Future Work

• Do the results extend to a larger population?
• What is the lifetime of data?
• What transformations are performed on the data?
• How effective can the proposed features be?

• Empirical study scheduled for November
Revealing the Copy and Paste Habits of End Users

*Kathryn T. Stolee, Sebastian Elbaum, and Gregg Rothermel
University of Nebraska-Lincoln

This work was supported in part by the EUSES Consortium through NSF-ITR 0324861 and 0325273, and CFDA#84.200A: Graduate Assistance in Areas of National Need (GAANN).