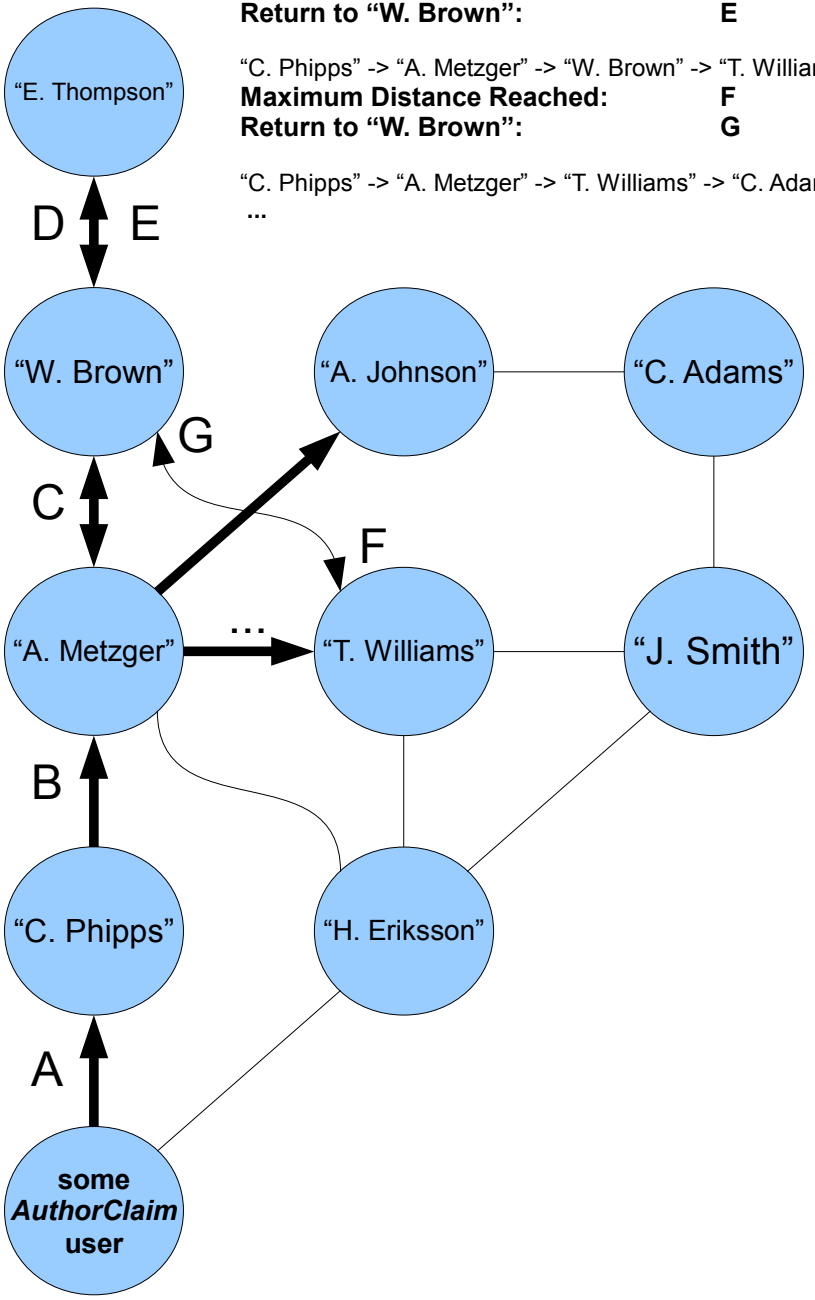
This is a working draft. Please see the AuthorProfile F. A. Q. at http://authorprofile.org/faq.txt

An Exploration into the Network (to a Distance of 4 Nodes):

"C. Phipps" > "A. Metzger" > "W. Brown" > "E. Thompson"
Maximum Distance Reached: D
Return to "W. Brown": E

"C. Phipps" -> "A. Metzger" -> "W. Brown" -> "T. Williams"
Maximum Distance Reached: F
Return to "W. Brown": G

"C. Phipps" -> "A. Metzger" -> "T. Williams" -> "C. Adams"
...



Citation network analysis is one of the primary operations of the system being developed for the *AuthorProfile* service. Each potentially identifiable author specified within the bibliographic metadata (i. e. "citation") records is rendered as a node within this social network. The relationships established through the intellectual collaboration between any given authors upon a given work (as specified within the *3lib* datasets) are rendered as network edges.

The possibility of resolving the possible identity* of each author specified within the *3lib* datasets would be impossible without the user profiles generated by *AuthorClaim*. Linking *AuthorClaim* users to one or more documents, they may also relate a given *AuthorClaim* user to several *author names*, identifying *author name variations*.

Currently, the citation network analysis performs a depth-first search of a set of trees. This set is limited to those trees for which the root node is an *AuthorClaim* user. Beginning with the root node, the immediate neighborhood of authors is explored by identifying those authors with whom this *AuthorClaim* user has directly collaborated upon any given work. The resident neighborhood for each node is then explored to a given depth within the network. The shortest path between each unresolved author related to an *AuthorClaim* user within this tree is represented in a data structure. This data structure is then serialized into a database.

It must be recalled that the time which elapses between updates to the *AuthorClaim* profiles and the next iteration for a network exploration focusing upon any *AuthorClaim* user is significant. That the structure of this network is constantly changing, and that the representations of the shortest paths are constantly inconsistent with its current state is a problem currently being addressed in the development of this system.

*As the process of resolving "author identities" local to this system relies upon human-curated data, the author hesitates to describe this process as "identification".

James R. Griffin III
02/28/12
AuthorProfile
Open Library Society