The Cascading Citations Analysis Project (C-CAP)

Dimitris A. Dervos (1), Nikolaos Samaras (2), Georgios Evangelidis (3), Theodore Folias (4), Y. Asmanidis (5)

(1,5) Information Technology Dept., Alexander Technology Educational Institute, Thessaloniki, Greece. Email: {dad, ypasm}@it.teithe.gr
(2,3,4) Applied Informatics Dept., University of Macedonia, Thessaloniki, Greece. Email: {samaras, gevan, folias}@uom.gr

1. The Cascading Citations Indexing Framework (c²IF)

The Cascading Citations Analysis Project (C-CAP) aims at increasing the granularity of the citation indexing paradigm in order to facilitate more detailed analysis. Citations are addressed at the (*article, author*) level. For one target (*article, author*) pair the direct as well the indirect citations received are considered up to a pre-specified '*depth*' value k (*1-gen*, *2-gen*, …, *k-gen* citations).
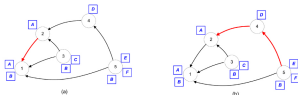

Figure 1. (a) one source: (2), two targets: (1,A), (1,B); one *1-gen* self-citation to (1,A), one *1-gen* citation to (1,B)
(b) one source: (5), one target: (2,A); one *2-gen* citation to (2,A)

A **chord** is an instance whereby a *k-gen* (k>1) citation is paired with an *1-gen* citation involving the same source (citing) article, and the same target (cited) (*article, author*) pair. The existence of a chord is taken to comprise an indication of increased value for the target, since it is cited both indirectly and directly by the corresponding source article.

Figure 2. One source: (5), two targets: (1,A), (1,B); one *3-chord* to (1,A), one self-*3-chord* to (1,B)

2. UAI_Sys/c²IF_Sys Pilot Implementation

To operate the cascading citations scheme, each author need be uniquely identified. In this respect, the C-CAP deliverables include a pilot universal author identifier (UAI) implementation.

Figure 3. UAI_Sys/c²IF_Sys architecture

3. The Testbed: ISI Science Citation Index Expanded

To test the system, six years of citations data (1999-2005) from the ISI Science Citation Index Expanded (ISI SCIE) were utilized. The dataset registers 7,364,211 research article records involving 165,822,522 citation instances. Following the data cleaning/preparation stage, 35,503,513 citation instances have been identified to satisfy the requirement that the cited articles are present in the dataset considered. Figure 4 presents the findings for 20 of the most highly cited authors taking into consideration citation results up to level 3, i.e. up to 3-gen citations, and 3-chords.

Figure 4. ISI SCIE (1999-2005) sample citation standings graphs