Hasan Davulcu Citation Statistics:

Citation indices

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Since 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Citations</strong></td>
<td>920</td>
<td>562</td>
</tr>
<tr>
<td><strong>h-index</strong></td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td><strong>i10-index</strong></td>
<td>22</td>
<td>16</td>
</tr>
</tbody>
</table>

Services Integration:

My highest cited papers in services integration published after 2002:


   [70 citations]:

   - **Web Service Interfaces**, Dirk Beyer (EPFL), Arindam Chakrabarti (Berkeley), Thomas A. Henzinger (Berkeley), WWW 2005:

     “CTR-S is a *formalism to model dynamic contracts between services.*”

   - **Computational Law**, Nathaniel Love (Stanford) and Michael Genesereth (Stanford) in International Conference on Artificial Intelligence and Law (2005):

     “*CTR-S representation is significantly more declarative and flexible than traditional business process/workflow representations. CTR-S also offers significant benefits over business process languages by easily integrating additional constraints into the definition of a process or workflow. One limitation of CTR-S is that it models one agent as the reasoner, seeking to augment the workflow with additional constraints, and all other agents in the world as a monolithic opponent— there is no unified view of multiple agents acting simultaneously in an environment.*”


     “*There has been noteworthy research in all modules mentioned for configuration ... Negotiation using game theory was discussed in CTR-S*”
Information Extraction and Web Integration:

My highest cited papers in information extraction and Web integration published after 2002:


[53 citations]:

- **Benchmarking natural-language parsers for biological applications using dependency graphs** Andrew B Clegg and Adrian J Shepherd in BMC Bioinformatics 2007:

  “Although much headway has been made using text processing methods based on linear pattern matching (e.g. regular expressions), the diversity and complexity of natural language has caused many researchers to integrate more sophisticated parsing methods into their biological NLP pipelines [IntEx]. This enables NLP systems to take into account the grammatical content of each sentence, including deeply nested structures, and dependencies between widely separated words or phrases that are hard to capture with superficial patterns.”

- **Multi-way relation classification: application to protein-protein interactions**, Barbara Rosario (Berkeley) and Marti A. Hearst (Berkeley) in Proceedings of the conference on Human Language Technology (2005):

  “In the BioNLP literature there have recently been a number of attempts to automatically extract protein-protein interactions from PubMed abstracts. Some approaches simply report that a relation exists between two proteins but do not determine which relation holds), while most others start with a list of interaction verbs and label only those sentences that contain these trigger words (IntEx).—Most of the existing methods also suffer from low recall because they use hand-built specialized templates or patterns. Some systems use link grammars in conjunction with trigger verbs instead of templates (IntEx).”

[27 citations]:


> “The Semantic Partitioner Algorithm traverses the DOM tree representation of a Web page in a top-down fashion to segment the content based on entropy. This work focuses on gathering and separating meta-data and their instances from various kinds of Web pages, which can empower the information retrieval.”


> “Future work involves extensive experimentation with the proposed focused crawling methods, as well as adapting some other Web page segmentation methods in the literature (e.g., the VIPS algorithm, which segments Web pages based on the visual clues, or the Semantic Partitioner method).”


[23 citations]:

- **AggregateRank: bringing order to web sites**, Guang Feng (Microsoft) et.al. in ACM SIGIR 2006

> “In the traditional view, there are two kinds of essential elements that make up of the Web. One is the web page and the other is the hyperlink, corresponding to the content and structure of the Web respectively. In recent years, many researchers have realized that the website, another element of the Web, has played a more and more important role in the Web search and mining applications [1][2][3][4][5]. As compared to the individual web page, the website can sometimes provide plenty of semantic information in the following aspects. First, pages in the same website may share the same IP, run on the same web server and database server, and be authored / maintained by the same person or organization. Second, there might be high correlations between pages in the same website, in terms of content, page layout and hyperlinks.”

[31 citations]:

- **Towards a Statistically Semantic Web**, Gerhard Weikum, Jens Graupmann, Ralf Schenkel, and Martin Theobald (Max-Planck Institute of Computer Science) in Conceptual Modeling – ER 2004

  “The first step in building an ontology is to create the nodes and edges. To this end, existing thesauri, lexicons, and other sources like geographic gazetteers (for names of countries, cities, rivers, etc. and their relationships) can be used. To further enhance the ontology, we crawled Web pages with HTML tables and forms, trying to extract relationships between table-header column and form-field names and the values in table cells and the pulldown menus of form fields. Such approaches are described in the literature (see, e.g., [OntoMiner]). Our experimental findings confirmed the potential value of these techniques, but also taught us that careful statistical thresholding is needed to eliminate noise and incorrect inferencing, once again a strong argument for the use of statistics.”

- **TaxaMiner: an experimentation framework for automated taxonomy bootstrapping**, Vipul Kashyap (Partners HealthCare) in IJWGS 2006

  “A complementary approach that uses the structure and content of HTML-based pages on the Web to generate ontologies is presented in [9].”

- **Identifying the Multiple Contexts of a Situation**, Aviv Segev (Technion - Israel Institute of Technology) in LNCS Volume 3946/2006 Modeling and Retrieval of Context.

  “Information seeking is the process in which people turn to information resources to increase their level of knowledge regarding their goals [4]. Although the basic concept of information seeking remains unchanged, the growing need for the automation of the process has called for innovative tools to assign some of the tasks involved in information seeking to the machine level. Thus, techniques for information seeking based on textual information are used, including the ontology tools Text-To-Onto [8], OntoMiner [6], and TexaMiner [7], to name a few, and databases are extensively used for the efficient storage and retrieval of information.”


“The realm of information science has produced an extensive body of literature and practice in ontology construction, e.g., [44]. Other undertakings, such as the DOGMA project [41], provide an engineering approach to ontology management. Work has been done in ontology learning, such as Text-To-Onto [24], Thematic Mapping [6], OntoMiner [7], and TexaMiner [15] to name a few. Finally, researchers in the field of knowledge representation have studied ontology interoperability, resulting in systems such as Chimaera [26] and Protégé [31].”


“Davulcu et al [6] used automated techniques for bootstrapping and populating specialized domain ontologies (concept taxonomies) by organizing and mining directories and links in a set of relevant web sites provided by the user.”