coDNA: A Visualization Tool for the Analysis of Peer Production

Ofer Arazy
University of Alberta
Alberta, Canada
University of Haifa, Israel
ofer.arazy@ualberta.ca

Henry Brausen
University of Alberta
Alberta, Canada
hbrausen@ualberta.ca

David Turner
University of Alberta
Alberta, Canada
dwt@ualberta.ca

Adam Balila
University of Haifa,
Haifa, Israel
adambalila@hotmail.com

Eleni Stroulia
University of Alberta,
Alberta, Canada
stroulia@ualberta.ca

Joel Lanir
University of Haifa,
Haifa, Israel
ylanir@is.haifa.ac.il

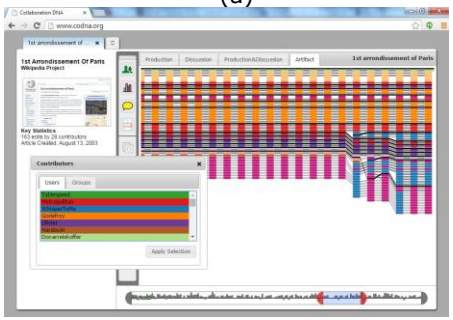
Abstract
Our demo is an information visualization tool designed to illustrate the temporal evolution of the peer production process. We combine comprehensive data extraction methods (automated, manual, machine learning) with user-friendly visualization techniques. Our visualization tool – *coDNA* – supports researchers in the development of grounded theory of peer production and allows practitioners to monitor production processes within their online community.

Introduction
Large scale collaborative efforts such as Wikipedia or open source software development projects represent a community-based model for the production of knowledge-based goods. In these peer-production projects, contributions of knowledge are made by volunteers, who self-organize to manage the production process.

The overarching objective of this project is to support the investigation of online production communities and enable the generation of grounded theory of peer production. In particular, we focus on the temporal dynamics of the process by which small product elements are contributed by volunteers and then integrated into a unified knowledge-based product, and we seek to delineate the sequential patterns of collaborative production.

The sheer scale and complexity of peer-production systems present a serious barrier to manual methods for identifying relevant patterns of behavior, thus calling for an automatic method for analyzing knowledge production processes. The availability of temporal data harvested from logs of IT systems supporting peer-production could be employed to track the interactions in socio-technical systems and capture the sequential contributions to a common artifact. Just

(a)



(b)

Screenshots of coDNA showing (a) 'Production' view with a filter on the type of activities and (b) 'Artefact' view with a filter on the contributors

as the Human Genome Project maps the sequences of genes in the human DNA, we chart the 'DNA' sequences of computer-mediated collaboration. We, thus, refer to our project as *Collaboration DNA*, or simply *coDNA*.

The coDNA visualization tool

Visualization makes relevant processes visible that would otherwise be difficult to interpret. This is of particular importance for scholars investigating online production communities and for the administrators and owners of these communities. To date, most of the visualizations have tended to focus on a single dimension presented a static view, capturing metrics such as total number of activities per contributor or the structure of relationships between contributors. We wish to go beyond current practices and develop a visualization tool that would capture the *temporal sequences of activities* in the evolution of peer production across *multiple dimensions*.

The software tool we will demo was developed to facilitate the exploration and analysis of 'collaboration DNA' sequences. The tool visualizes data collected from a series of peer production projects. Several design principles guided the development of *coDNA* (http://codna.org), as described below:

- **Temporal dynamics**. Given our focus on the temporal evolution of peer production, it is essential that we capture the timing of events and that the visualization emphasizes temporal aspects.
- **Nested organizations**. We view online production communities as nested organizations, where each organization operates multiple projects. For example, Wikipedia is an organization that operates multiple projects – each project being the collaborative authoring on an encyclopedic entry.
- **Project vs. Contributor focus**. When studying collaboration patterns we may focus on a particular project, mapping all participants' activities in this project. Alternatively, we may want to focus on the activities of a single contributor across multiple projects. Our tool supports both views.
- **Production & Coordination**. Activities in online production communities include the contribution of product elements (e.g. editing a Wikipedia page), as well as coordination activities (i.e. discussions around task allocation). The tool supports both activity kinds.
- **Activities**. The basic building blocks of 'coDNA' are activities, specifically production and coordination activities. We capture and represent key attributes of those activities such as: date & time, activity type, scope of activity, and the contributor.
- **Contributors and roles**. Contributors play different roles in the organization, often moving between roles. The visualization tool records and presents the organization role of a contributor.
- **Internal and external events**. Both internal (e.g. community's decision to resolve a conflict) or external (change in the competitive landscape) events may impact collaboration patterns. Hence, we seek to record and visualize both event types.
- **Process vs. Product focus**. The outcome of a peer production process in a knowledge-based product is an artefact. We capture and represent the architecture of this product, as well as the relationships between contributions of small elements and the product's architecture.