

iChase: Supporting Exploration and Awareness of Editing Activities on Wikipedia

Nathalie Henry Riche
Microsoft Research
Redmond WA 98052, USA
nath@microsoft.com

Bongshin Lee
Microsoft Research
Redmond WA 98052, USA
bongshin@microsoft.com

Fanny Chevalier
Microsoft-INRIA joint center
LRI, Université Paris Sud, France
Fanny.Chevalier@inria.fr

ABSTRACT

To increase its credibility and preserve the trust of its readers, Wikipedia needs to ensure a good quality of its articles. To that end, it is critical for Wikipedia administrators to be aware of contributors' editing activity to monitor vandalism, encourage reliable contributors to work on specific articles, or find mentors for new contributors. In this paper, we present iChase, a novel interactive visualization tool to provide administrators with better awareness of editing activities on Wikipedia. Unlike the currently used visualizations that provide only page-centric information, iChase visualizes the trend of activities for two entity types, articles and contributors. iChase is based on two heatmaps (one for each entity type) synchronized to one timeline. It allows users to interactively explore the history of changes by drilling down into specific articles and contributors, or time points to access the details of the changes. We also present a case study to illustrate how iChase can be used to monitor editing activities of Wikipedia authors, as well as a usability study. We conclude by discussing the strengths and weaknesses of iChase.

Categories and Subject Descriptors

H.5.2 [Information Systems]: User Interfaces — User-centered design;

General Terms

Design.

Keywords

Wikipedia visualization, timeline visualization, interaction.

1. INTRODUCTION

Since Ward Cunningham developed the first wiki [1] in 1994, wikis have been astonishingly successful, especially when used as collaborative knowledge building tools [2][1]. A famous example is Wikipedia, a free online collaborative encyclopedia consisting of millions of pages created by hundreds of thousands users in hundreds of languages. Since late 2004, several French contributors have been working towards releasing an offline version of Wikipedia to be distributed on DVD: Wikipedia 1.0 [4]. To make this first release of Wikipedia possible, a number of active contributors volunteered to administrate projects (sets of articles on a given topic) on core topics and ensure that key

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AVI'10, May 25–29, 2010, Roma, Italy.

Copyright 2010 ACM 1-58113-000-0/00/0004...\$5.00.

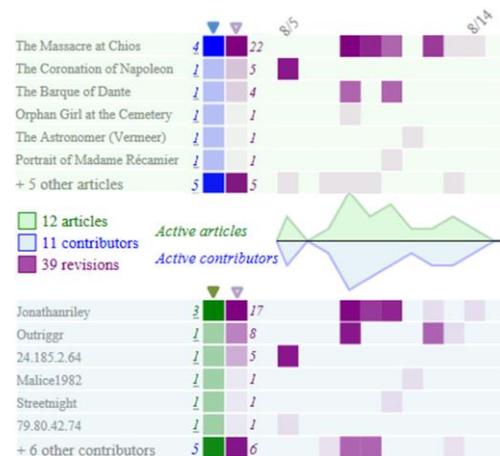


Figure 1. iChase in monitoring mode of one week activity on the WikiProject Louvre Paintings

articles reach a good level of completeness and quality. To better monitor each project, administrators established a set of measures to rank the quality, importance and progress of each article [5]. However, even with these measures, maintaining an awareness of the activity of a project with dozens of articles and hundreds of contributors is a challenging task.

Wikipedia's great success has raised interests from many information visualization researchers. For example, History flow [6], WikipediaViz [7], and JWikiVis [8] visualize the editing activity of given articles and depict their evolution. A couple of visualizations such as Chromograms [9] show the editing activities of contributors, attempting to characterize different profiles. Surprisingly, only a few visualizations such as WikiDashboard [10] focused on providing day-to-day awareness of the activity of articles and contributors. While WikiDashboard has been proven helpful to occasional contributors and readers to assess the status of individual articles[11], it fails to support the current needs of Wikipedia 1.0's administrators: since administrators need to monitor multiple articles and contributors at the same time, they are required to navigate from page to page to assess the most recent activities on the project (from both the articles' and contributors' perspective). Today, to maintain awareness of a project activity, these administrators' common practice is to review the textual list of revisions since they last logged in. They report this practice to be not only tedious but also difficult for them to maintain a mental map of the evolution of their projects and to decide where the effort should be directed.

To help administrators acquire a better awareness of the activity of their projects, we designed and developed iChase (Figure 1). iChase is an interactive multi-scale visualization of the editing

activity of a Wikipedia project. It provides a general overview of the project activity to textual revisions and access to articles. The core contribution of iChase is to provide a dual-entity exploration for both articles and contributors in a single visualization. It is designed to ease back and forth navigation between articles and contributors, a crucial interaction for a number of scenarios. For example, an administrator identifies a surge of activities on one article, she explores the activities of each contributor for this article, validating most of the revisions, but detecting a potentially harmful contributor. The administrator can review the activities of this contributor, access each article edited by this user while keeping the article she was initially reviewing in view and further pursue her exploration.

In this paper, we first review the related work and describe the results of a participatory design workshop we ran with 16 administrators of Wikipedia 1.0. We describe our design goals and present iChase's user interface. We demonstrate how iChase fits our design goals through a case study and present the results of a usability study with 4 administrators and 4 occasional contributors. We conclude with the strengths and weakness of iChase and future work.

2. RELATED WORK

2.1 Visualizations of Editing Activities

There have been several efforts on visualizing editing histories on Wikipedia pages. Both History flow [6][12] and JWikiVis [8] visualize trends in revision histories by showing relationships between multiple versions of a page and the general evolution of its content. These visualizations are primarily page-centric and do not provide any awareness on the contributors' editing patterns.

WikiDashboard [10] and WikipediaViz [7] provide embedded visualizations of articles or users activity within each Wikipedia page. Their goal is to increase social transparency and help Wikipedia users to get a better awareness of the editing activities on pages and contributors. However, users have to navigate through multiple pages to understand editing patterns for a set of pages or a set of contributors. Moreover, these visualizations do not allow changes in activity granularity: WikiDashboard only shows daily activity of articles and user while WikipediaViz always visualizes the whole history of an article.

Chromograms [9] is a more general visualization attempting to characterize the activity trends of highly active contributors by encoding the editing histories for a collection of pages into color bars through a simple color coding scheme. However, rather than providing awareness on contributors' activity to Wikipedia users, Chromograms is a researcher tool, aiming at discovering styles and general rhythms of editing activity on Wikipedia.

In summary, existing visualizations either provide awareness to regular Wikipedia users, reading or editing individual Wikipedia pages or focus on analyzing the global activity on Wikipedia. To our knowledge, there is no visualization supporting our administrators' needs: the exploration and monitoring of a set of pages and a set of contributors.

2.2 Visualizations for Bipartite Datasets

The Wikipedia data we wish to explore, contributors and articles, linked by revisions, constitutes a bipartite graph evolving over time. Many datasets are bipartite, i.e., contain two main entity types. Examples include papers written by authors, keywords included in documents or photos tagged by people. Traditional

(bipartite) network visualizations (e.g., node-link diagrams [13]) are not designed to represent links carrying a large amount of information such as a list of textual revisions. Moreover, while there have been several attempts at visualizing network evolving over time [14][15], it still remains a challenge to provide users with an overview of the evolution and details of the changes.

Tackling these complex datasets from a different perspective, NetLens proposes a content-actor data model and visualizes the two entity types in symmetric coordinated views [16]. NetLens allows users to explore the data by iteratively refining queries and interacting with visual overviews and sorted lists in either view. While it provides similar features to iChase, there are several differences. First, NetLens uses an individual bar graph per entity element (e.g., for the Wikipedia data, one bar graph per article and one per contributor). While this fine grained level of details gives more explicit trends on each element, it requires significantly more screen estate, making NetLens not scale well. Second, NetLens provides iterative query refining by sending filtered elements from one view to the other. Although an interaction history is provided, it remains difficult to keep track of the current view content and of the exploration process. Finally, NetLens does not provide time exploration or multiple levels of details.

2.3 Focus+Context Table Visualization

TableLens first applied the Focus+Context technique to visualize large tabular data [17]. In addition to text representation, it used bar graphs for the quantitative variables and colored rectangle for categorical variables. Line Graph Explorer [18] extended TableLens to support line graph data. LiveRAC is another system visualizing a large collection of time-series data organized as a table using Focus+Context technique for exploration [1]. In LiveRAC, rows and columns respectively represent devices and sets of device parameters. Neither rows nor columns are ordered according to time. Instead, the time visualization is represented inside each cell of the table. Thus, each cell contains a multiple line graph representing the evolution of a set of parameter values for a given device. The strength of LiveRAC is to provide a smooth transition from overview to details using the accordion drawing metaphor. While we use a similar Focus+Context technique in iChase, the design philosophy is different. iChase is centered on a unique timeline, each heatmap showing the activities of a set of elements (articles or contributors) over time rather than depicting their correlations with each other. We believe this perspective is better suited to provide general awareness of activity evolving over time.

3. Participatory design workshop

To better understand the needs of administrators, we organized a one full day participatory design workshop with 16 members of Wikipedia 1.0. During the workshop, we collected their scenarios, discussed their current practice, introduced them to a wide variety of visualizations, and organized brainstorming sessions and video prototyping of their ideal system. From our discussion and the video prototypes, we extracted a set of 4 main tasks and 7 design goals for an interactive visualization.

3.1 Tasks and scenarios

To maintain awareness of the activity of a project, our participants currently parse the last revisions and rely on their knowledge of the contributors, articles and activity rhythm of the project to detect unusual activity. Following is a short scenario explaining their current exploratory practice.

On Monday, John logs in to monitor his project: “Oh, there was an editing activity on the LouisXIV article on Sunday. It is unusual as Mary, the main contributor on this article, usually does not work on Sunday. Who has made this edit? Hmm... It is a contributor named marc29, someone I do not know. What did he change? This revision content does not look right. Let’s see what other articles marc29 has edited. Two more on the same day: Napoleon and LouisXV. Several contributors I know are working on LouisXV; let’s see if they reverted the changes by marc29. Hmm... It does not look like someone has been editing this article after marc29. I wonder if they are aware of these changes. Let’s see, this past week, George has been editing a lot LouisXV, I should send him and Mary an email asking them to check marc29’s edits.”

A large number of scenarios were collected and our discussions with administrators confirm the exploratory nature of their current practice and the serendipitous discoveries that help them build an awareness of the activity. While this makes it difficult to describe their strategy of exploration, we identified four main tasks.

(T1) Detection of unusual activity and tracking of vandalism: this task concerns day-to-day monitoring of the project. Project administrators’ time is mostly spent in patrolling: detecting unusual activity and tracking vandalism. They control the content of new articles, of long or numerous revisions on a single article or from a single user, as well as revisions from new or unknown users. They consider this task very important as vandalism affects the credibility of Wikipedia.

(T2) Awareness of articles’ status: to make Wikipedia 1.0 happen, administrators need to monitor the progress of all project’s articles and to identify the ones requiring more effort.

(T3) Awareness of contributors’ interests: project administrators need to be aware of contributor’s interests to encourage them to contribute on specific articles or mentor new contributors. They also need to monitor the activity of new contributors.

(T4) Overview of project activity after a long period of absence: our participants reported that it was difficult for them to catch up with the overall activity of their projects after a long absence (e.g., a few weeks long vacation). They reported feeling overwhelmed by the amount of data to review to come back on track with the current state of the project when using existing watchlists.

3.2 General Design Goals

From the 5 video prototypes created by our participants and a number of comments, we formulated 7 design goals for iChase.

(G1) Getting an overview of the overall activity of both articles and contributors: providing a dual-focus overview of both articles’ and contributors’ activity to quickly assess the project activity for a given period of time.

(G2) Exploring revisions on a given article: showing the history of the activity for each article, including the list of contributors who worked on it and the text revisions.

(G3) Exploring revisions made by a given contributor: showing the history of the activity of each contributor, including the list of revisions of the articles she edited. In general, providing awareness of (new) contributors’ activity and therefore getting to know their interests is important to invite them to work on specific articles or to collaborate with other contributors.

(G4) Accessing text revisions, whole articles’ and contributors’ details: providing detailed information and content of articles and revisions; providing information about contributors.

(G5) Identifying salient activity (e.g., outlier detection): supporting the identification of unusual, salient activity on the project is important to track vandalism and control the content of articles and revisions. Examples include excessive activity of a particular contributor on a particular article or numerous revisions.

(G6) Supporting iterative queries: back and forth navigation between articles and contributors. We named iterative queries the navigation our participants perform when successively exploring the activity of articles and contributors. An example scenario using iterative queries is provided in section 3.1.

(G7) Maximizing the recall: when monitoring dozens of articles and contributors, it is difficult to review all the changes and quickly assess the project’s progress when logging in after a rather long period of time. The system should provide a stable visualization, maximizing the recall between each session.

4. iCHASE

The majority of participants in the workshop led towards augmenting existing Wikipedia pages with interactive visualization. While it would preserve their current practice, augmenting existing article and contributor pages (as do WikipediaViz [7] and WikiDashboard [10]) fails to support several goals (G5, G6 and G7). For example, administrators are required to navigate from pages to pages to gain awareness of what contributors have been doing or which articles have been edited for a given period of time. This makes the exploration process tedious as administrators need to remember some information while they are navigating the set of pages. It is also hard for them to keep track of their exploration paths. To avoid this tedious page to page navigation, we designed iChase as a stand-alone application.

4.1 User Interface

iChase (Figure 2) is built upon an augmented timeline designed to show the activity of two coupled entities at several levels of details. The novelty of iChase lies in the design of this dual-entity timeline and the interactions to navigate within and between two entity types. We considered the following three principles when designing iChase:

1. Simple and compact visual representation to limit the learning phase and provide a glanceable visualization.
2. Interactions to support iterative queries (G6) and to investigate specific activities of both contributors and articles.
3. Scalability to cover the wide range of the number of articles and contributors administrators monitor as well as the time duration they wish to investigate.

4.1.1 Visualization

iChase displays three distinct entities: articles, contributors and revisions (connections between articles and contributors). To help users easily identify each entity within all visual representations, iChase uses a consistent color-coding: green, blue, and purple represent articles, contributors, and revisions respectively.



Figure 2. iChase showing the WikiProject *French Revolution* (12 months).

The time range and granularity of the data visualized are shown in (1). In iChase time is represented horizontally; timestamps indicated at regular intervals (2). The first heatmap (3) represents the set of articles monitored in this project including the total number of contributors and revisions for each article (4). A legend shows the number of articles, contributors and revisions active for the given time range (5). Two line graphs indicate the evolution of the number of articles and contributors active over time (6). The second heatmap contains an expanded row (7) representing a given contributor and the two articles he has edited. The rest of the heatmap is collapsed to provide context (8). A summary is provided for collapsed contributors on mouse over (9).

Timeline (Figures 2(1) and 2(2)). In iChase, time flows from left to right and each column represents a particular point in time. Users can set the start and end dates of the data to be visualized and select the granularity of the visualization.

Activity heatmaps (Figure 2(3)). Each entity type is represented by a heatmap in which rows are individual elements (articles or contributors) and columns are time periods. The cells of these heatmaps are color-coded by the number of revisions that occurred for the corresponding element on a particular time period. These heatmap representations convey general trends about a large quantity of data not by massively aggregating or summarizing them but by representing as many data items as possible in cells potentially very small but varying in color intensity. Therefore, heatmaps convey general project activity awareness while displaying activity at the almost finest level of details. By default, articles are ordered by their creation date. However, users can reorder rows by number of active articles (contributors) or number of revisions.

iChase supports two strategies to handle large heatmaps: one for exploration and the other for monitoring. To ensure access to all the data for exploration, iChase offers a scrolling panel for each heatmap. To help users monitor the activity of a project and provide everyday-awareness with minimal interaction, iChase allows users to fit the heatmap in a given view by letting them manipulate the height of rows that are not focused, to show more or less entities; or to aggregate rows that cannot fit in the view (Figure 1). This solution provides a summary of the project activity in a single screen, however there are obvious tradeoffs

when aggregating items and summarizing activities. Indeed, it might be difficult to detect outliers and irregularities when using aggregated rows.

Aggregated row indicators (Figure 2(4)). iChase provides additional visual representations to summarize activity in each row to help users detect and compare the editing activity of articles or contributors. For example, for the article row, the two rectangles in each row represent the number of active contributors and the total number of revisions, respectively, for a specific period of time. To remain consistent, iChase color-coded these cells according to their entity type. iChase also shows actual values to ease comparison as color intensities might be difficult to differentiate for small differences.

Activity line graphs (Figure 2(6)). The two line graphs placed opposite to each other show the number of edited articles and active contributors, respectively, over time. They help users understand the general trends and rhythms of editing activities over time.

Legend (Figure 2(5)). The legend summarizes the total number of articles, contributors, and revisions currently shown in iChase. It also serves as a reminder of the color coding of each entity.

4.1.2 Interactions

Mouse over (Figure 2(9)). iChase offers visual and textual feedback when users place the mouse cursor over visual elements. For example, when users move the cursor over a purple cell, iChase highlights both the column of the corresponding time point and the row of the corresponding article or contributor. In

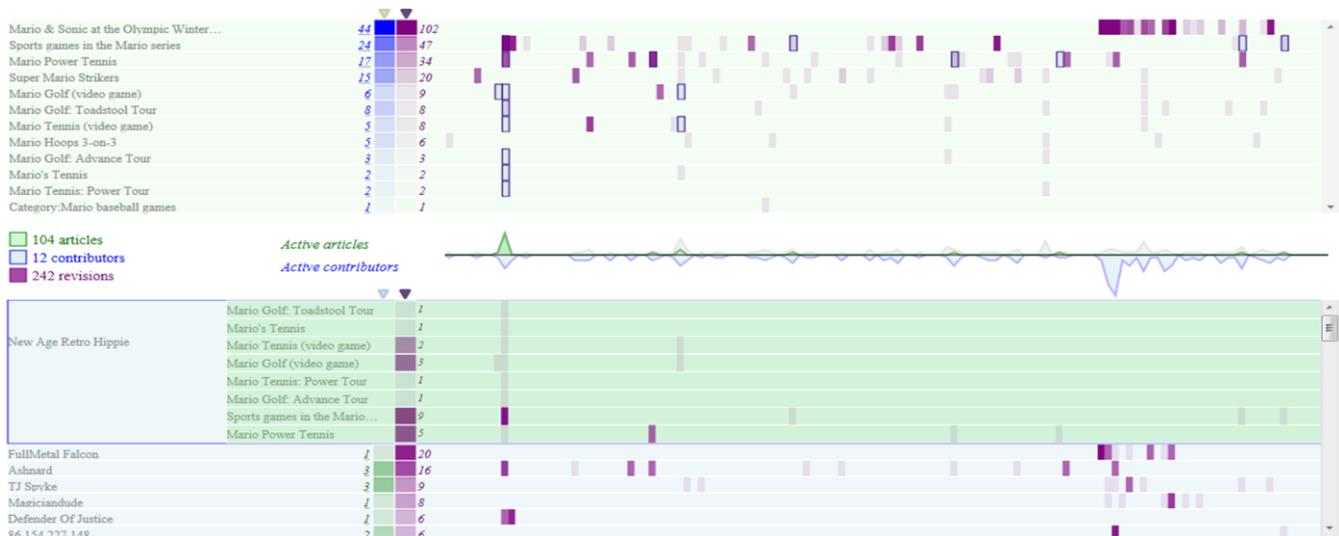


Figure 3. Activity Rhythms of the WikiProject *Mario Sport Games* (4 months)

addition, iChase provides a tooltip showing the detail information (the time, the number of edited articles or active contributors, and the number of revisions the cell contains). When users move the cursor over the line graphs, iChase highlights the column of the corresponding time.

Drilling into time. iChase visualizes editing activity over a consecutive time range at different granularities. It provides interactions to drill down into time and display more details about these activities using a Focus+Context technique. Clicking on a column (either over the column header of the heatmaps or the activity line graphs) expands the corresponding time range by stretching the column to occupy the available screen width, and displays more information on the revisions.

Rather than allowing users to alter the content of Wikipedia pages through iChase, we chose to provide links towards pages. Clicking on one of these links opens the Wikipedia page showing the difference between the original article content and its content after the revision in the default web browser. From this web page, users can reach the article content and its history of revisions if desired.

Drilling into entities (Figures 2(7) and 2(8)). iChase supports exploration of the activity of a given article or contributor more in details. When users click on the corresponding row header, the selected row is expanded and the rest of the heatmap rows are collapsed into a minimum height. We decided to collapse non-selected rows instead of filtering them out to provide the user with contextual information. For example, the collapsed rows can convey the general trend of activity even with 1 pixel height. Placing the mouse cursor over a collapsed row shows a tooltip feedback, thus helping users decide whether or not to select the collapsed row. When expanding a row (e.g., an article), iChase displays more details about its activity by breaking down the row into several sub-rows representing the activity of each contributor on the selected article. When an article is in focus, the blue line graph is filtered and shows the number of contributors editing this article during that time span. In addition, the cells containing

revisions on this article are highlighted with a green border in the contributors' heatmap. Similarly, when a contributor is in focus, the green line graph is filtered, showing the number of articles he has edited during that time span, and corresponding cells in the articles' heatmap are highlighted.

To support back and forth navigation (iterative queries), iChase provides an interaction on the sub-rows. When users click on the sub-row (e.g., a contributor of the focused article), iChase emphasizes the row (representing the selected contributor) in the other heatmap by highlighting its border. To further help users identify the focused row, iChase also fades out other rows. In addition, iChase expands the row to show the articles this particular contributor has worked on. When users click on one of the articles this contributor has changed in the second heatmap, iChase collapses the previously selected article (in the first heatmap) and expands the new corresponding article.

Collapsing rows and columns. As described above, iChase offers a two-level exploration: users can drill down into time (columns) and/or into rows (articles or contributors). Thus, users can reach the lowest level of a cell in two distinct ways (first expand time, then article/contributor or the inverse). To provide a consistent behavior, iChase uses left click on rows and columns to expand them. Left click on a cell expands the row and column together. Right click on rows and columns collapses them. Right click on a cell collapses the corresponding row and column at once.

Animations. To help users understand the change of view when expanding a row or column, we provide animations inspired by accordion drawing [18].

4.2 Implementation and Performance

iChase prototype is implemented in C# using the Windows Presentation Foundation (WPF) framework. iChase can handle thousands of revisions on hundreds of articles and contributors on a standard Pentium 4 PC. When disabling animations, iChase can handle more than 30,000 revisions in interactive response time.

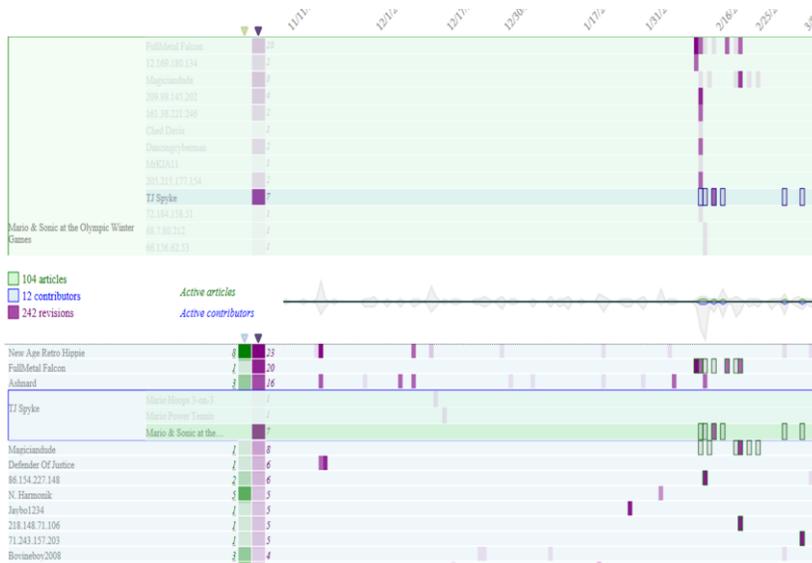


Figure 4. Activity Rhythms of *Mario Sports Games* (4 months)

5. CASE STUDY

We demonstrate how iChase supports our design goals using 4 WikiProjects: *Louvre paintings* (Figure 1), *French Revolution* (Figure 2), *Mario Sports Games* (Figures 3, 4, and 6) and *Winter Olympics* (Figure 5).

5.1 Overall Activity (G1)

The activity line graphs, the repartition of the cells in the heatmaps and their color intensities are the primary elements depicting the overall activity of a project and help administrators catch up after a long period of absence (T4). For example, the line graphs in Figure 1 show that there was no activity on 8/6 and that the most active period was two days after. The overall symmetric shape of the activity lines indicates that the editions were mostly done on a 1:1 basis, i.e., generally one article was edited by one contributor. By glancing at the articles heatmap ordered by number of active contributors, the administrator can assess that *Massacre at Chios* was the main focus of this week's activity while the users heatmap reveals that *Jonathanriley* was the most active contributor. Figures 2 and 3 show two additional examples. Figure 2 depicts a regular activity (mostly with a 1:1 pattern) whereas Figure 3 shows a burst of activity in the last third of the period, in which many users have been editing a few articles. Heatmaps in both pictures reveal a different everyday activity. The *French Revolution* project (Figure 2) shows activity on many articles, whereas only two were active in *Mario Sports Games* (Figure 3). In addition to the activity line graphs, legend giving the raw numbers of active contributors, articles and revisions, help to detect unusual activity.

5.2 Exploring Revisions for an Article (G2)

By observing the article heatmap, users can detect a burst of activity and investigate the rhythm of edits of a specific article. For example, the first article of *Mario Sports Games* (Figure 3) suddenly shows a burst of activity: a succession of horizontally aligned dark squares after none activity. It actually corresponds to the creation of the article. By expanding this article, one can see revisions from multiple contributors each day (Figure 4). Figure 4 shows that only three contributors edited repeatedly over time while the others were punctual writers. In addition, the line graph

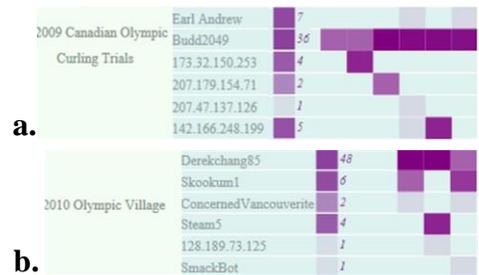


Figure 5. Details on a salient activity of two articles from *Winter Olympics*

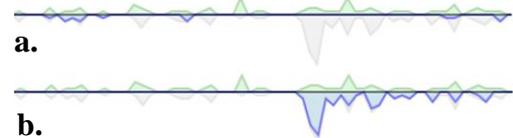


Figure 6. Filtered activity graphs of (a) the first and (b) the second articles of Figure 3.

is filtered and shows the activity of the contributors for the selected article (Figure 6(a)). One can immediately identify that the burst of contributor activity on the project was caused by this first article. As a comparison, Figure 6(b) shows the activity of the second article, exhibiting a more regular pattern. Finally, zooming in time and reviewing the revisions' details reveals that content was added mostly on sections "characters" and "development" during the first three days. Then information was suppressed by lack of reliable sources and finally, typos and formatting were fixed. This type of information helps administrators assess what are the states of given articles and their creation history (T2).

5.3 Exploring Contributors' Revisions (G3)

Glancing at the heatmap of the contributors of *Mario Sports Games* (Figure 4) reveals two distinct rhythms of activity for the two most active contributors. The second contributor edited a single article in a short period of time and gradually became less active. This is visually conveyed by a succession of dark squares. The first contributor (*New Age Retro Hottie*) has a much more regular activity as showed by the dashed pattern and worked on multiple articles. By expanding this contributor (Figure 3), one can notice that he mostly edited one or two articles at a time, except for a particular day where he edited 8 articles together. Zooming into the revisions for this day reveals that he created a general page on *Mario Sports Games* and added a link towards it in each individual article.

5.4 Accessing Text Revisions (G4)

Some patterns detected in the heatmap can draw attention to particular revisions. For example, several contributors editing an article repeatedly in a short period of time may indicate that they have successfully collaborated or, on the contrary, that they disagreed on a particular section. The presence of vertically aligned dark squares on the same article depicts this salient activity (Figure 5). To identify if the activity rhythm of these contributors is fortuitous, positive or negative, one needs to access the comments and text of the revisions. On the *Olympic village* article (Figure 5(b)) for example, by clicking on the successive revisions to access the details, one can identify that both *Dereckchang85* and *Skookum1* edited at the same time period

twice. Accessing text revisions shows that *Dereckchang85* is adding content to the article while *Skookum1* is fixing the formatting template. Further review of the revisions reveals a warning message from *Skookum1* ordering *Dereckchang85* to stop altering the template. This information is important since new contributors may feel rejected or offended by more expert contributors. By noticing this conflict, the administrator may contact *Dereckchang85* and suggest him an appropriate mentor (T3) to keep him involved in the project.

5.5 Identifying Salient Activities (G5)

iChase attracts administrator's attention on salient activity (T1) in multiple ways. First, by looking at the activity line graphs, administrators can detect unusual or irregular activity patterns such as drop or peak of activities. A peak in the articles line indicates that many articles have been edited at that time. If the contributor line does not reflect this peak and shows a very small amount of active contributors, it may suggest suspicious or repetitive activity such as contributors or robots fixing format and typos. The burst of activity of the line graph of *Mario Sports Games* (Figure 3) is worth investigating. By clicking on the top most active article, the line graph shows that the burst is due to this article (Figure 6(a)). This information shows individual activity rhythms and helps detect salient patterns. As administrators have their own knowledge of the regular project activity rhythms and external information on the project contributors (such as vacation time), they can detect irregularity or outliers on the heatmaps. We identified a few patterns indicating a potential irregular activity (Figure 5). They mostly consist in several contributors editing in a short period of time possibly repeatedly. As administrators have previous knowledge of the contributors' activity, identifying the name of a given contributor in these patterns may also help them quickly assess the situation.

5.6 Supporting iterative Queries (G6)

iChase supports iterative queries by providing two different perspectives on the data (articles' perspective and contributors' perspective) and a simple mechanism linking them. Salient activity is the entry point of iterative queries; iChase provides interaction to track the suspicious activity of this contributor. iChase supports the scenario described in 3.1 in a couple of clicks.

5.7 Stable Visualization for Monitoring (G7)

To maximize the recall when logging multiple times and provide a stable visualization, iChase maintains the order of heatmap rows from session to session. To help monitoring, we also envision adding visual clues on the activity evolution (by adding an up and down arrow to show the article activity since the last logging).

6. USER STUDY

We collected feedback on iChase from 8 participants: 4 administrators and 4 occasional contributors. After presenting the visual encoding and interaction available in iChase, we asked our participants to analyze two datasets of our case study (*Winter Olympics* and *Mario Sports Games*). The administrators were also provided with one month activity of their own watchlist as an additional dataset. To collect maximum insights without over guiding our participants, we divided the study into three phases. First, we asked the participants to describe any insight that they found in the data. Second, we asked more guided but still high-level questions (e.g., compare the editing rhythm of two given contributors). Finally, we asked specific questions (e.g., find

contributors working on the same two articles) in order to cover all of iChase's features. During the session, we asked our participants to think aloud: 1) insights they found in the data, 2) describe their strategy and 3) point out usability issues. In addition, administrators had to fill out a questionnaire comparing iChase to their current monitoring tool that mainly consist of textual list of revisions.

6.1 Insights found with iChase

We collected about 50 insights from these datasets prior to the study. We made a note each time the participants found one of the 50 listed items. All 8 participants discovered them all, and mainly during the first two phases (without experimenters asking specific questions). All of the administrators agreed that iChase allows them to keep track of contributors and articles activity (T2 and T3), possibly after a long period of absence (T4), and help them detecting abnormal or suspicious activities (T1).

All administrators commented that it was especially easy to detect potential vandalism using iChase. For example, they quickly identified contributors editing multiple articles in a short period of time: *"Either this user has updated the articles (e.g., updated the formatting) or a bot made these changes or this is an act of vandalism. [With iChase,] I can check the comments and then revoke all the contributions of this user in case he is a vandal."* By exploring the activity of his own project, one administrator was able to identify a new contributor and quickly assess his interests (T3): *"This unknown contributor is interested in the same topics as I am! I would like to contact him, either to share sources, to synchronize or to discuss... With the current watchlist, I did not even notice this user has edited several times articles of my project. I would have had to wait for him to become a regular contributor."* One administrator explained that he sometimes acted as an external reviewer to help solve conflicts in other projects. He commented on how iChase could help him assess the overall activity of an unfamiliar project: *"With iChase, I can understand the overall situation at a glance: who and what articles are concerned, when the conflict started, ..."*

6.2 Usability

As many new visualization tools, iChase requires some learning time. It took occasional contributors the first half of the study (around 30 min) to learn to decode the heatmaps and discover the tight coupling between these. Only one occasional contributor did not take advantage of the iterative queries. However, she could still answer all the questions. The administrators were particularly fast in learning iChase (around 15 min) and used all of its functionalities. One of them was provided with the tool without any explanations and was able to use and describe all the features after only 10 minutes of practice. This reveals that iChase can be used by a wide audience with a reasonable learning time.

Overall, our participants reported a small number of minor usability problems. The major one concerned the right click to close rows and/or columns. Almost all of them commented that they got lost when closing a cell, and that they expected closing their last opened item (row or column) instead of closing both row and column. We agree and will fix this issue.

6.3 Comparison with existing tools

Administrators participating in the study all used textual lists of revisions (a tool called LiveRC[20] or watchlist). We asked them to rate iChase and their current tools for four aspects using a Likert scale from 1 to 5 (see Table 1). iChase received higher

scores than their current tool in all aspects. When asked if they would use iChase to replace their current tools, all 4 administrators answered that they probably would but only with real time feeds (our study was done with archived files). Finally, they commented on the usability of iChase: “*iChase is pleasant to use and to look at. The information is visually represented, which is easy to use compare to text lists.*”

	iChase	Current tools
Overall project awareness	4.7 (0.6)	1.7 (0.6)
Exploration of articles’ activity	4.7 (0.6)	2.7 (0.6)
Exploration of contributors’ activity	4.3 (0.6)	2.3 (1.5)
Support for iterative queries	4.7 (0.6)	3.0 (1.0)

Table 1: Administrators ratings from 1 (worst) to 5 (best) for iChase and their current tools. Average (Standard Deviation).

6.4 Missing features

All the experts asked for more information on the type of contributor (e.g., bot, IP address, registered, administrator, etc) and to mark revisions identified as revert. They also requested shortcuts to perform administrators’ actions: revert a revision, block a contributor, access a discussion page, etc. Concerning the interaction, administrators were mostly interested in marking and filtering heatmap rows. They wanted to color regular, registered contributors (as supported by their current tools) and possibly hide them from the view. They also wanted to filter out articles already reviewed. Several administrators commented that filtering would allow them to refine their awareness of the project activity. They explained that they would probably be aware of highly active articles, but that filtering them out would help them assess the activity on the rest of the articles, which are more difficult to follow. Several participants also asked for more advanced sorting features, such as sorting by total length of revisions for a specific period. Overall, they did not consider these features essential for using iChase but commented on their additional benefits.

7. CONCLUSION AND FUTURE WORK

We presented a novel visualization called iChase designed and developed to provide Wikipedia administrators with better awareness of article editing activities. Unlike other visualizations that provide only page-centric information, iChase visualizes the trend of activities for both articles and contributors by using two heatmaps synchronized to one activity timeline. iChase makes it possible for administrators to interactively explore the history of changes by drilling into specific articles and contributors, or time points to access the details of the changes. To show the utility of iChase, we presented a case study and a usability study collecting feedback from administrators and occasional contributors.

In the future, we are planning to iterate on iChase to fix the minor usability issues raised by our participants and add the missing features identified by the administrators during our initial study. We plan to run longitudinal studies *in situ* with Wikipedia administrators to better understand their needs and assess how they use iChase. We also plan to organize a second participatory design workshop focusing on team awareness in order to provide WikiProjects members and administrators with tools to communicate and collaborate more effectively.

8. REFERENCES

[1] WikiWikiWeb, <http://en.wikipedia.org/wiki/WikiWikiWeb>
 [2] FoxPro Wiki, <http://fox.wikis.com>
 [3] MeatballWiki, <http://www.usemod.com/cgi-bin/mb.pl>
 [4] Wikipedia: 1.0 Editorial Team., http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team
 [5] Wikipedia 1.0/Core topics, http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Core_topics
 [6] F.B. Viégas, M. Wattenberg, K. Dave. Studying cooperation and conflict between authors with history flow visualizations, Proc. Conference on Human Factors in Computing Systems (CHI '04), pp. 575-582, 2004.
 [7] F. Chevalier, S. Huot and J.D. Fekete WikipediaViz: Conveying Article Quality for Casual Wikipedia Readers, Proc. PacificVis, 2010.
 [8] J. Gawryjolek, P. Gawrysiak, The Analysis and Visualization of Entries in Wiki Services, Advances in Intelligent Web Mastering, vol. 43, pp. 118-123, 2007.
 [9] M. Wattenberg, F.B. Viégas, K.J. Hollenbach, Visualizing Activity on Wikipedia with Chromograms, Proc. INTERACT'07, LNCS, vol. 4663, pp. 272-287, 2007.
 [10] B. Suh, E.H. Chi, A. Kittur, B.A. Pendleton, Lifting the veil: improving accountability and social transparency in Wikipedia with wikidashboard, Proc. Conf. on Human Factors in Computing Systems (CHI '08), 1037-1040, 2008.
 [11] P. Pirolli, E. Wollny, B. Suh, So you know you’re getting the best possible information: a tool that increases wikipedia credibility, Proc. Conf. on Human Factors in Computing Systems (CHI'09), pp. 1505-1508, 2009.
 [12] F.B. Viegas, M. Wattenberg, J. Kriss, F. van Ham, Talk Before You Type: Coordination in Wikipedia, Hawaii International Conference on System Sciences, pp. 78a, 2007.
 [13] I. Herman, G. Melançon, and M. S. Marshall, Graph Visualization and Navigation in Information Visualization: a Survey, IEEE Trans. Visualization and Computer Graphics, vol. 6, no. 1, pp. 24-43, 2000.
 [14] C. Erten, P. J. Harding, S. G. Kobourov, K. Wampler, G. Yee, GraphAEL: Graph Animations with Evolving Layouts, Proc. Graph Drawing, pp. 98-110, 2003.
 [15] M. Ghoniem, H. Cambazard, J.-D. Fekete, N. Jussien, Peeking in solver strategies using explanations visualization of dynamic graphs for constraint programming, Proc. SoftVis, pp. 27-36, 2005.
 [16] H. Kang, C. Plaisant, B. Lee, B.B. Bederson, NetLens: Iterative Exploration of Content-Actor Network Data, Information Visualization Journal, vol. 6, pp. 18-31, 2007.
 [17] R. Rao, S.K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information, Proc. Conf. on Human Factors in Computing Systems (CHI '94), 318-322, 1994.
 [18] R. Kincaid, H. Lam, Line graph explorer: scalable display of line graphs using Focus+Context, Proc. Advanced Visual Interfaces (AVI '06), pp. 404-411, 2006.
 [19] P. McLachlan, T. Munzner, E. Koutsofios, S. North, LiveRAC, Interactive Visual Exploration of System Management Time-Series Data, Proc. Conf. on Human Factors in Computing Systems (CHI '08), 1483-1492, 2008.
 [20] LiveRC (Live Recent Changes), <http://en.wikipedia.org/wiki/User:EDUCA33E/LiveRC>