Clickstream Logs

A clickstream log provides information about user interactions (e.g., clicks, page loads, scrolls, etc.) with a web interface. Clickstream data can be used to construct a detailed picture of an individual user’s behavior and engagement, and can also be used to examine trends in learner behavior.

Eventing data structure

Recorded events will include a subset of the following data fields. The event key will always be included; other fields are dependent on the exact nature of the event.

The structure and content of Coursera’s eventing data changes as platform features and services evolve. The events below are subject to modification or removal. Other events not documented here may also appear in some or all clickstream exports; data related to these additional events should not be assumed to be reliable or consistent.

|  |  |
| --- | --- |
| key | type of user action being recorded |
| value | records any specialized information specific to the particular event |
| username | anonymized *session\_user\_id* of the user engaging in the interaction |
| timestamp | time that the event reached the Coursera server (UTC timestamp in milliseconds) |
| page\_url | URL of the page on which the event was recorded |
| client | component of Coursera serving the content with which the user is interacting |
| session | cookie-stored value used to track individual client machines (note that this is different from a browsing session) |
| language | list of ISO language codes used when determining which language to serve the site in |
| from | page that the user navigated to this one from (reported by the user’s browser) |
| user\_ip | the user’s IP address (may list multiple proxy IP addresses) |
| user\_agent | User Agent string provided by the user’s browser |
| [12] screen | serialized JSON array, indicating the pixel dimensions of the user’s screen |

**Event details**

Two common events merit further explanation; these are documented in detail below.

**“pageview” events**

Pageview events are recorded when a user accesses a page within a course experience (i.e., pages with a URL of the form class.coursera.org/sessionname-###) on Coursera.

Pageviews are not recorded when a user accesses central Coursera pages (e.g., the course browser, course dashboard, or course description pages), or when a user opens a page to view a file hosted on Coursera’s servers (e.g., lecture notes uploaded as a course asset). Pageviews are also not recorded for external (non-Coursera) pages, and Coursera does not track clickthrough rates to external pages, even if those pages are accessed from links in a course site.

**“user.video.lecture.action” events**

Video lecture events are recorded for a variety of interactions involving course videos. These events are currently surfaced by a third party plugin for the media engine powering the video player; thus, there may be inconsistencies in the data, and Coursera cannot guarantee the quality or reliability of video action event logs.

Video lecture events include several specialized data fields. The most important of these for most purposes is the event **type**, which provides more details on the recorded interaction.

The type field usually lists one of the following interactions:

|  |  |
| --- | --- |
| play | the video player started playing; indicates that a user opened a video or resumed play following a pause or completion of an in-video quiz |
| pause | the video player stopped playing; indicates that a user selected “pause”, encountered an in-video quiz, or reached the end of the video |
| ratechange | a user changed the playback speed of the video |
| seeked | a user clicked on the playback bar to jump to a particular location in the video (may be called automatically after certain error events) |
| stalled | the user’s browser is having trouble downloading the data it needs to continue playing, but has not yet encountered a de facto error |
| error | the video player encountered an error; the type of error will be recorded |

In addition to the type field, video lecture events also include the following special fields:

|  |  |
| --- | --- |
| currentTime | playback position of the video player (seconds) |
| playbackRate | current playback speed, as a multiplier against the standard rate of 1.0 |
| paused | whether the video player was paused at the time the event occurred |
| error | error code indicating that an error occurred; error=0 indicates that no error occurred, this value should be 0 |
| networkState | state of the network (as detected by the video player) at the time the event occurred |
| readyState | preparedness state of the video player at the time the event occurred |
| eventTimestamp | standard UTC timestamp (milliseconds) |
| initTimestamp | time the event being tracked was initiated (may be unreliable) |
| prevTime | playback time recorded before the event (may be unreliable) |

**Analyzing clickstream logs**

There are two common approaches to analyzing clickstream logs. The first is to convert the data to a relational format for use in traditional database software. This approach can be very useful for a particular subset of common queries against the data (i.e., “I want to see all of this user’s video interactions in order”); however, the large size of the data may strain the default memory and time constraints imposed by the software.

Programmatic approaches are more effective for questions that are difficult to answer with relational analysis (e.g., “How long did a user spend watching a video?”). [MapReduce](http://en.wikipedia.org/wiki/MapReduce) is one example of a program that can be applied to this type of problem.

# Data Sharing and Research Policies

Click the link below to download Coursera's Data Sharing Policy as and Research Policy as PDF files.

* [DataSharingPolicy.pdf](https://partner.coursera.help/hc/en-us/article_attachments/201755839/DataSharingPolicy.pdf) (400 KB)
* [CourseraResearchPolicy (2).pdf](https://partner.coursera.help/hc/en-us/article_attachments/202245969/CourseraResearchPolicy__2_.pdf) (300 KB)