



Making course content accessible

Problems



Institution



Instructor



Student

Institutional problems



No insight into how institution is doing



Hard to identify what to focus on



Lawsuits because of legal requirements

Instructor problems



Lack of awareness of what they should be doing

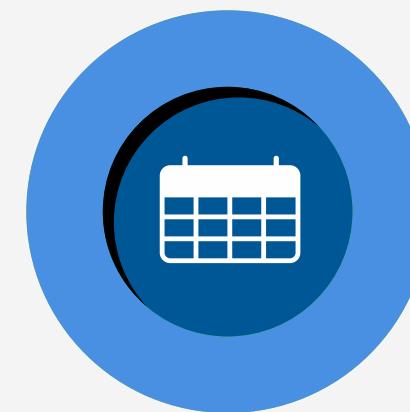


Lack of guidance on how to improve accessibility

Student problems



Explicit alternative
format requests



Long delays on receiving
requested alternative formats



Excludes students without
recognized disabilities



Learning Management System

- The Learning Management System is an important player in this
- Planning to provide integration with all major Learning Management Systems



Workflow

- Instructor adds course content to course site

The screenshot shows a course management system interface. On the left is a dark sidebar with white icons and text links: Home, Announcements, Assignments, Discussions, Grades, People, Pages, **Files** (which is highlighted with a blue background), Calendar, Syllabus, Inbox, Quizzes, Modules, and Settings. To the right of the sidebar is a main content area. At the top right, it says "BIO101 > Files". Below that is a search bar with the placeholder "Search for files" and a magnifying glass icon. Under the search bar is a tree view of files: "Biology 101" is expanded, showing "Midterm" and "unfiled". To the right of the tree view is a list of files with small thumbnail icons: "Const..." (image), "Gonzal..." (document), "Midterm" (folder), "NRC-E..." (image), "playing" (document), "unfiled" (folder), and "Zenlive" (document). At the bottom of the main content area is a progress bar showing "0% of 524.3 MB used".

BIO101 > Files

Search for files

Home
Announcements
Assignments
Discussions
Grades
People
Pages
Files
Calendar
Syllabus
Inbox
Quizzes
Modules
Settings

Biology 101

Midterm

unfiled

Name ▲

Const...

Gonzal...

Midterm

NRC-E...

playing

unfiled

Zenlive

0% of 524.3 MB used

Automated accessibility checklist



Automated accessibility checklist based
on content type

Machine Learning Algorithms

- Full structural and visual analysis to learn semantics of document
- Identify headings, heading structure, paragraphs, footers, tables, lists, mathematical formulas, etc.

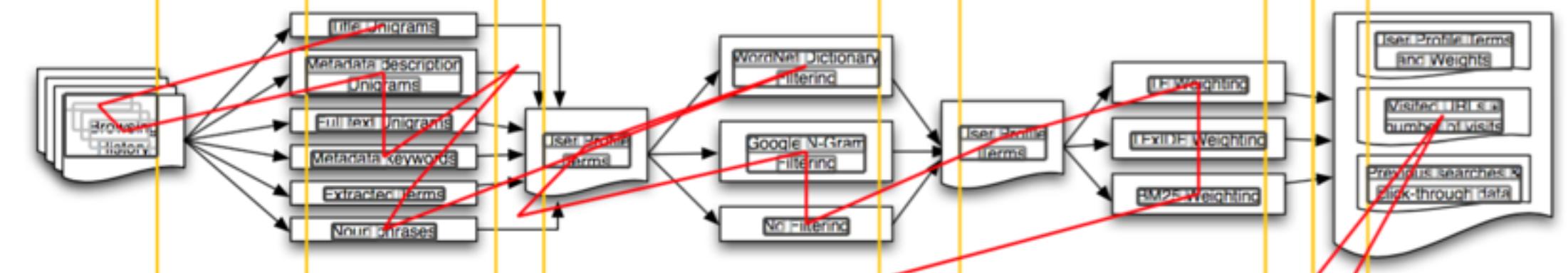


Figure 1: User Profile Generation Steps and Workflow

Metric	Total	Min	Max	Mean
Page Visits	530,334	51	53,459	10,607
Unique Page Visits	218,228	36	26,756	4,365
Google Searches	39,838	1	4,203	797
Bing Searches	186	1	53	4
Yahoo Searches	87	1	29	2
Wikipedia Pages	1,728	0	235	35

Every time a user leaves a non-secure (non-https) web page, the add-on transmits the user's unique identifier, the page URL, the visit duration, the current date and time, and the length of the source HTML to the server. The server then attempts to fetch the source HTML of this page. This is performed server-side to ensure that only publicly-visible data is used. Once the source HTML is received, the server compares its length to the length received from AlterEgo. If the length difference is smaller than 50 characters, the HTML is accepted and saved along with the other page visit data. Otherwise, we assume the content probably came from a password protected but non-secure site (e.g. Facebook, Hotmail, etc.) and the record is discarded.

Participants for this study were recruited via a website explaining the purpose and consequences to potential users, publicized on various e-mail lists, resulting in 50 participants taking part. Whilst we expect that most of these participants are employed in the IT industry due to the recruitment process, a number of people outside of the IT industry without significant web search experience participated as well. The add-on captured data for three months from March to May 2010. As shown in Table 1, a total of 530,334 page visits (or an average of 10,607 page visits per user) were recorded. 58% of the visits were to unique pages. The add-on also recorded 39,838 Google searches, 186 Bing searches and 87 Yahoo! searches, indicating that our users were strongly biased towards Google as their search engine, hence Google was used as the baseline in our experiments. An average user issued 797 queries over the three months, indicating that at least 7.5% of all non-secure web requests were search related.

3.1.2 Data Extraction

We considered the following summaries of the content viewed by users in building the user profile:

Full Text Unigrams

Table 2: Extracted terms from the AlterEgo website and the Wikipedia page about Mallorca

AlterEgo	Mallorca
add-ons	majorca
Nicolaas	palma
Matthijs	island
CSTIT	spanish
Nicolaas Matthijs	balearic
Language Processing	cathedral
Cambridge	Palma de Mallorca
keyword extraction	port

Title Unigrams

The words inside any `<title>` tag on the html pages.

Metadata Description Unigrams

The content inside any `<meta name="description">` tag

Metadata Keywords Unigrams

The content inside any `<meta name="keywords">` tag

Extracted Terms

We implemented the Term Extraction algorithm as presented in [31], running it on the full text of each visited web page. It attempts to summarize the web page's text into a set of important keywords. This algorithm uses the C/NC method, which uses a combination of linguistic and statistical information to score each term. Term candidates are found using a number of linguistic patterns and are assigned a weight based on the frequency of the term and its subterms. This is supplemented with term re-extraction using the Viterbi algorithm. The outcome of this algorithm run on two sample web pages can be seen in Table 2.

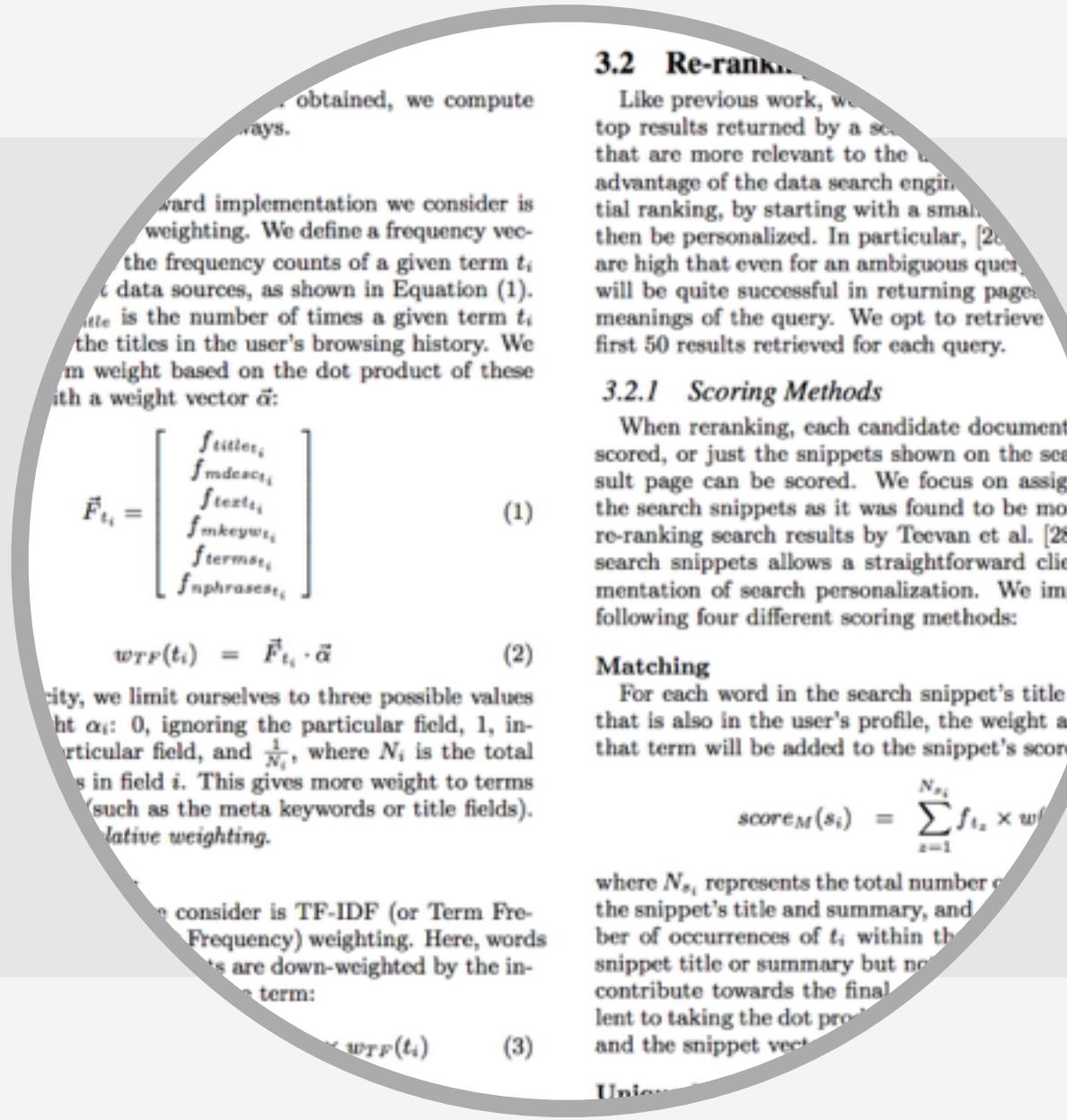
Noun Phrases

Noun phrases were extracted by taking the text from each web page and splitting it into sentences using a sentence splitter from the OpenNLP Tools³. The OpenNLP tokenization script was then run on each sentence. The tokenized sentences were tagged using the Clark & Curran Statistical Language Parser⁴ [3], which assigns a constituent tree to the sentence and part of speech tags to each word. Noun phrases were then extracted from this constituent tree.

3.1.3 Term List Filtering

To reduce the number of noisy terms in our user representation, we also tried filtering terms by removing infrequent words or words not in WordNet. However, neither of these were found to be beneficial. Therefore we do not discuss

Alternative Accessible Formats



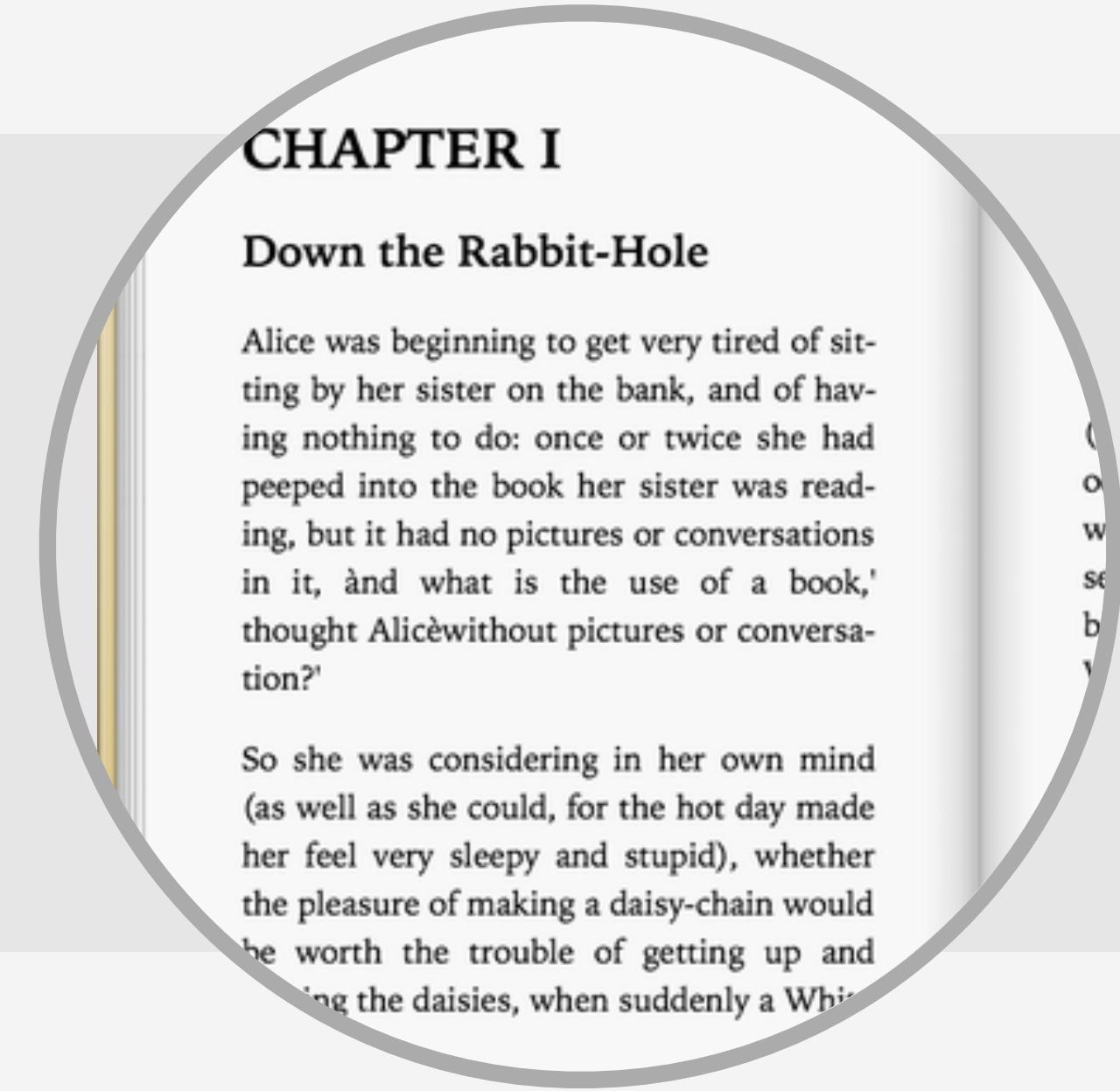
Enhance original

Automatically inject fixes into the original document



HTML

High quality semantic HTML version of the content

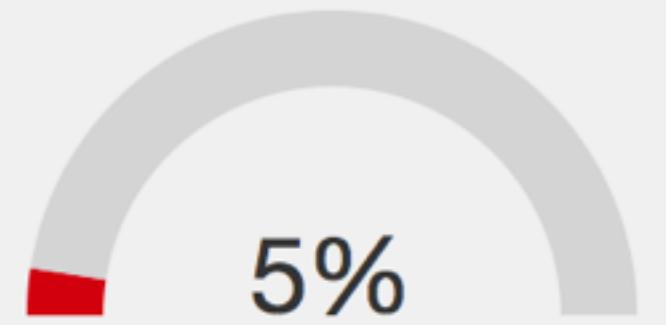


Other alternative formats

ePub, audio, electronic braille, etc.

Instructor feedback

- Provide guidance to instructors on how they can improve accessibility of their content
- Aims to generate change in behavior over time



This PDF has a poor accessibility score and can be difficult to use for a large number of students. Make this PDF much more usable by making a few improvements.



PDF

Tag this PDF

Lift the accessibility score of this document to at least **30%** by exporting it as a tagged PDF

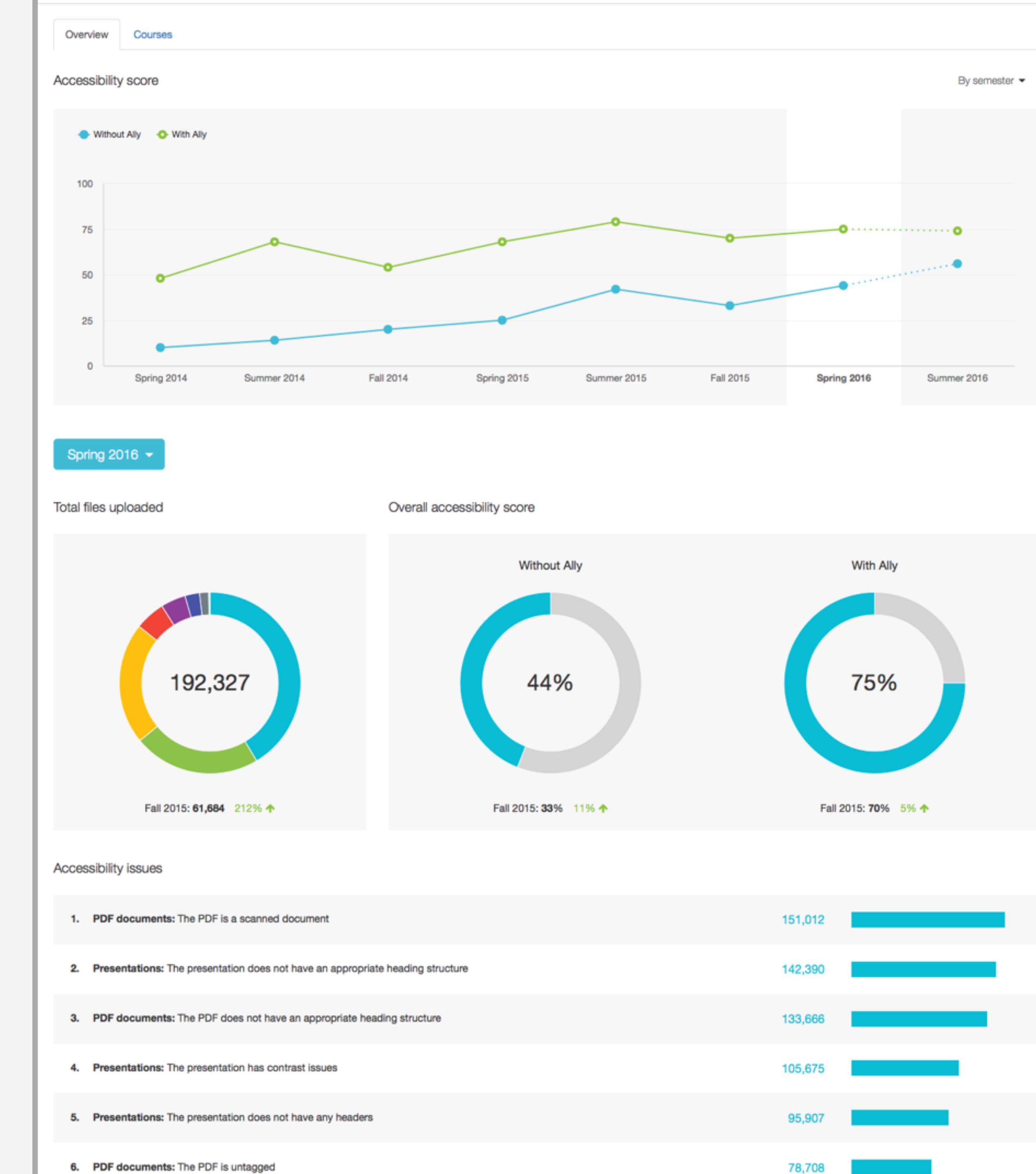


Show me how



Institutional report

- Provide baseline to institution on how they're doing
- Helps identify where problem areas are, what to focus on, who to target, etc.



Demo

The image displays a tablet and a smartphone side-by-side, illustrating a mobile application interface for managing accessibility scores.

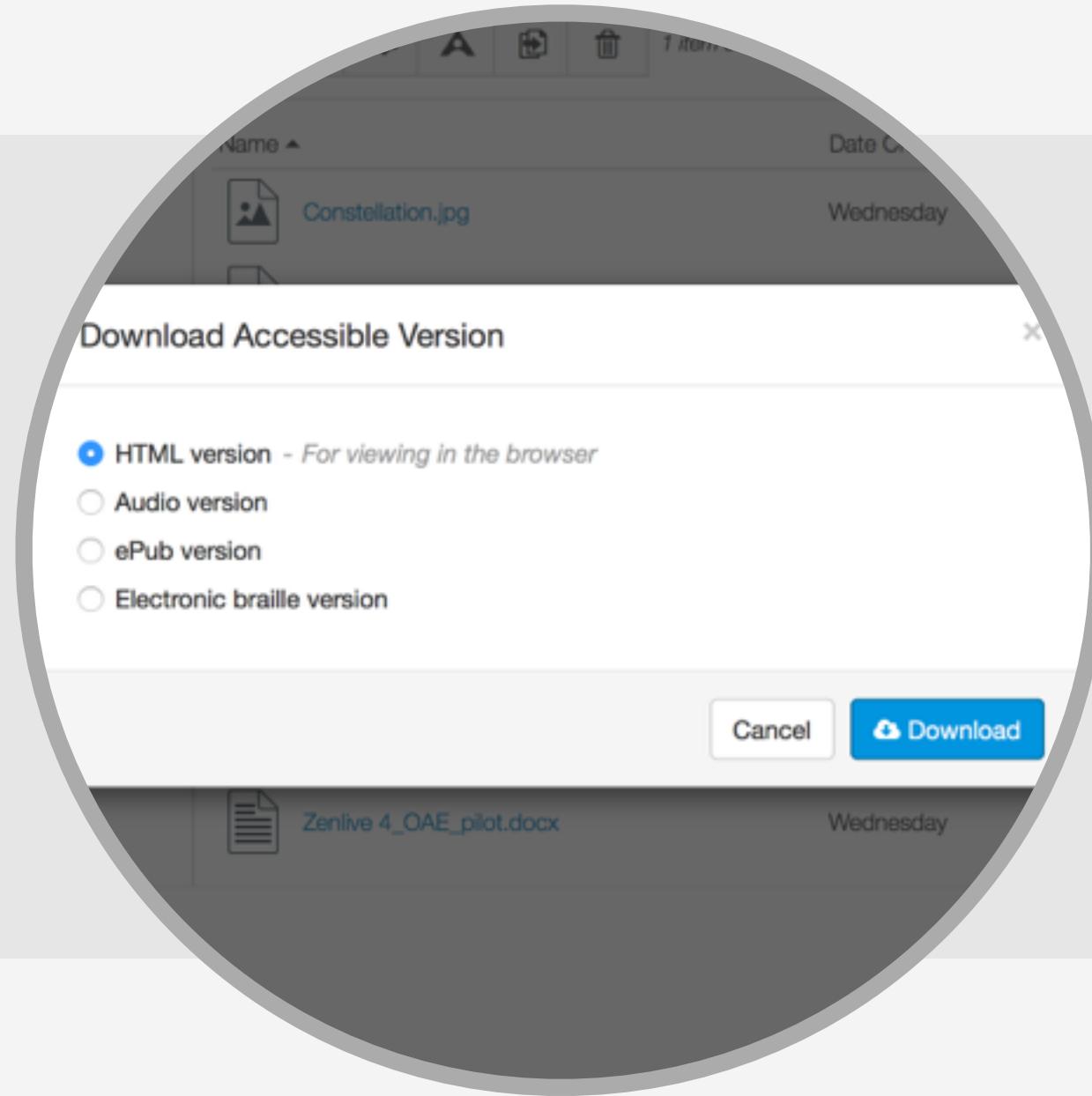
Tablet Dashboard:

- Left Sidebar:** Includes icons for Dashboard, Courses, Admin, Calendar, and Info.
- Top Navigation:** Shows "Overview" and "Courses".
- Section: Accessibility score**
 - Legend:** "Without Ally" (blue circle) and "With Ally" (green circle).
 - Line Graph:** Compares accessibility scores over time (Spring 2014 to Summer 2016). The "With Ally" series (green) starts at approximately 68% in Spring 2014 and rises to about 75% by Summer 2016. The "Without Ally" series (blue) starts at approximately 60% in Spring 2014 and rises to about 72% by Summer 2016.
 - Filter:** "By semester ▾" and a dropdown menu showing "2016 ▾".
- Section: Overall accessibility score**
 - Without Ally:** A donut chart showing 192,327 documents with a 44% overall accessibility score.
 - With Ally:** A donut chart showing 75% overall accessibility score.

Smartphone Detail View:

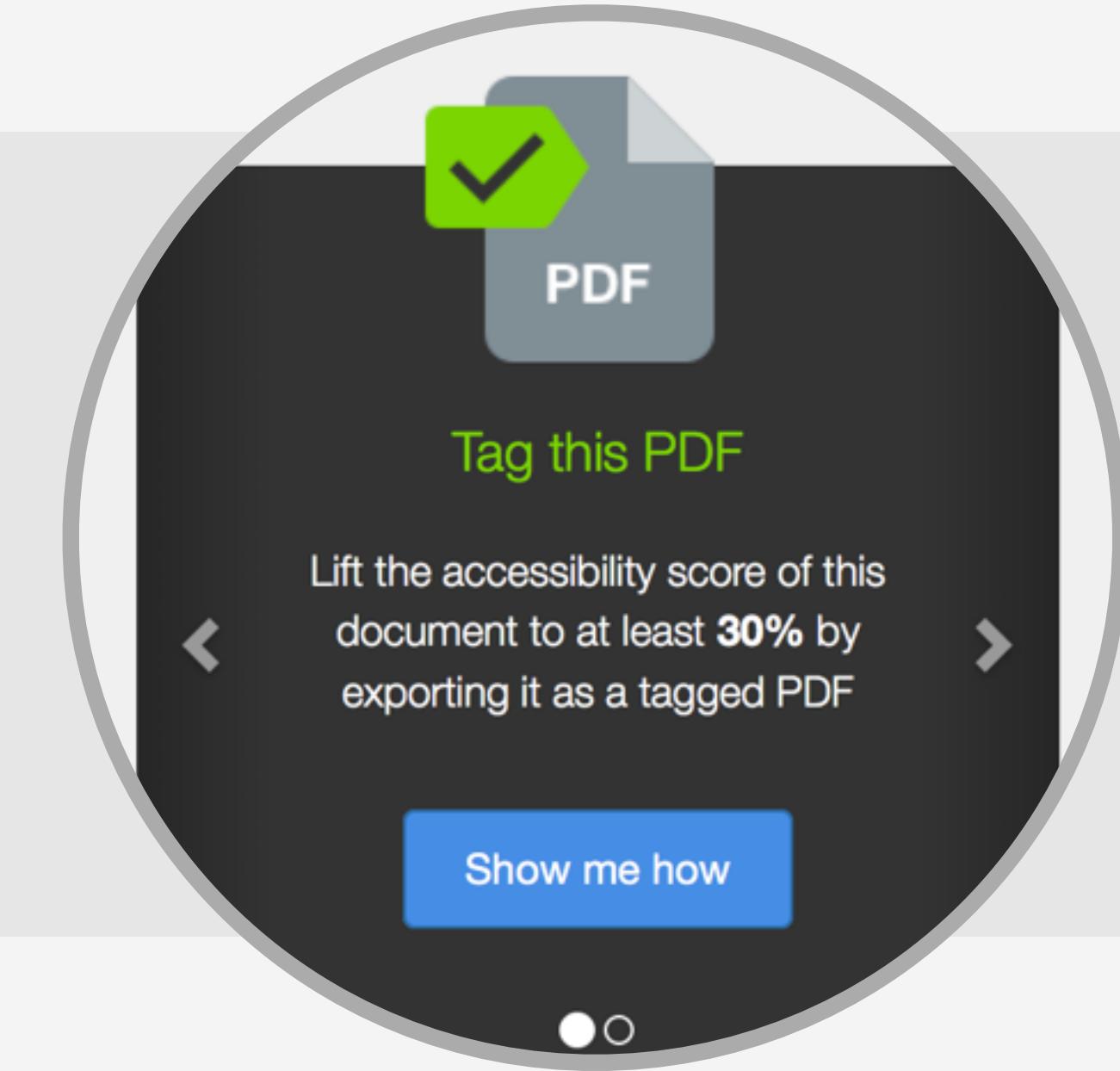
- Title:** Accessibility score for **Personalizing Web Search using Long Term Browsing History.pdf**
- Score:** 5%
- Description:** This PDF has a poor accessibility score and can be difficult to use for a large number of students. Make this PDF much more usable by making a few improvements.
- Icon:** A PDF file icon with a green checkmark.
- Action:** **Tag this PDF**
- Text:** Lift the accessibility score of this document to at least 30% by exporting it as a tagged PDF
- Button:** **Show me how**

Ally



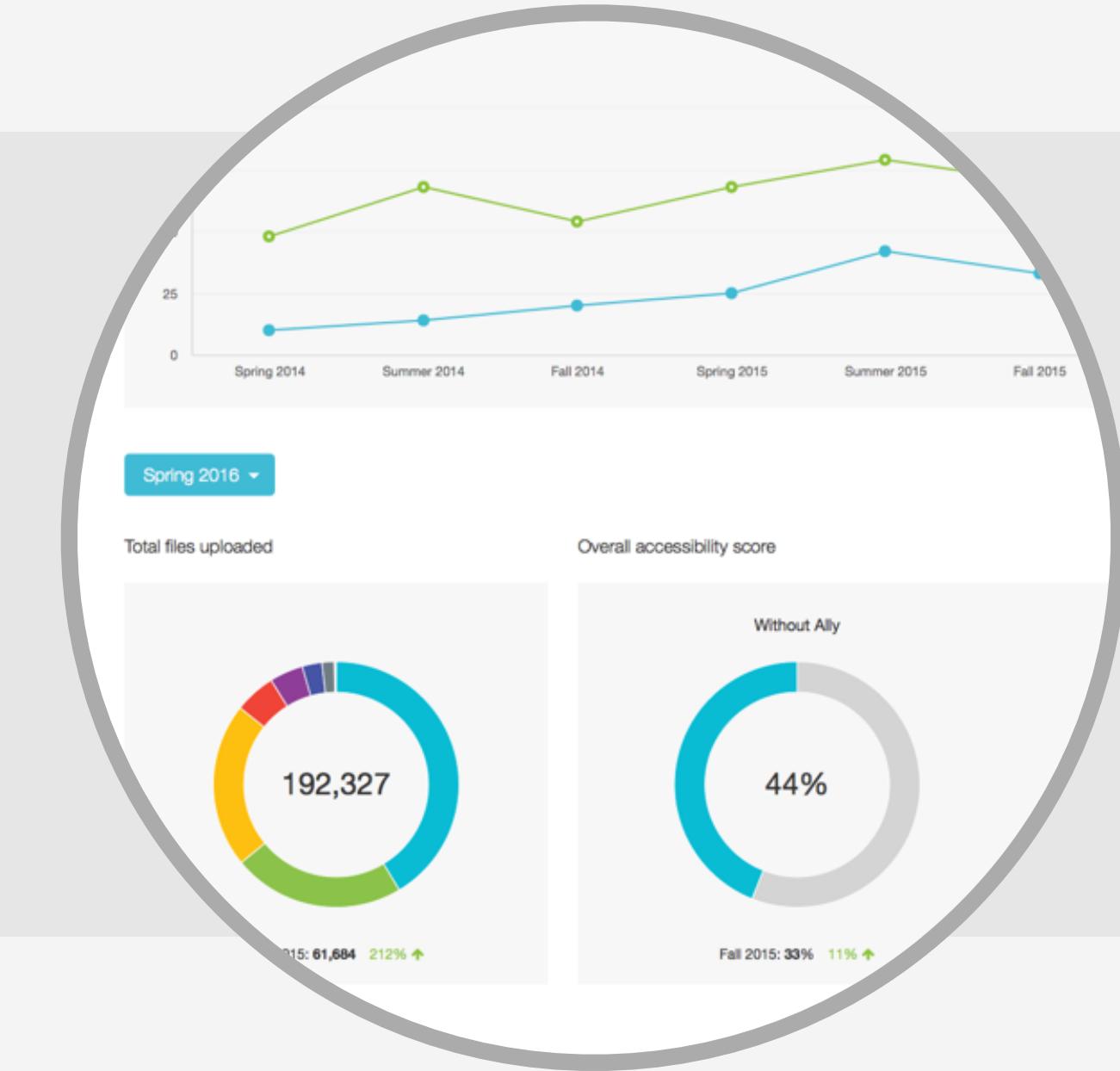
Alternative accessible formats

Automatically check for accessibility issues and generate alternative accessible formats



Instructor feedback

Guide instructors on how to improve course content accessible and alter future behavior



Institutional report

Provides ammunition to further improve course content accessibility at the institution

Interested ?



We are looking for
institutions to pilot Ally



We are looking for
individuals to join the
Ally evaluation panel

Nicolaas Matthijs



info@ally.ac