Using Turnitin to contextualize citations & dig into big textual data

Serge Noiret
EUI History Information Specialist (Ph.D.)

Library Research Skills Session

Wednesday 7 March 2018, 13.30-15.00, Sala del Capitolo, Badia Fiesolana
Library Research Skills Sessions

- Searching, citing and managing documentation with Zotero & Endnote
  Find out more...

- Suring the legal semantic web
  Find out more...

- Endnote training session
  Find out more...

- Advanced Zotero for lawyers
  Find out more...

- Research Data: how to prepare and submit datasets for preservation and sharing in the new

- Using Turnitin to contextualize citations, perform distant reading and dig into big textual

- Zotero advanced features
  Find out more...

- EU legislation and documentation on the web
  Find out more...

- Full-text resources for social and political scientists
  Find out more...
The EUI Dean of Studies and the Academic Service decided to introduce systematically the use of anti-plagiarism software. The reason was for single Ph.D. researchers to look at the various chapters and drafts of their dissertation during the four years research/writing process and verify the originality of the contents. But going beyond plagiarism is possible digging into Turnitin’s concept of similarity search. During a Library mission in 2013, I proposed to deliver a paper called «Beyond Turnitin», at the RRCHNM, Fairfax, Virginia, 2013. The idea was that using Turnitin, you were able to verify the validity of your own writings, quotations and bibliographies but also, to discover who else was working in your field of research so to extend your knowledge of available literature. With Turnitin you will discover who is quoting the same source and in which context.
Beyond Turnitin and anti-plagiarism softwares

BY SERGE NOIRET

In my university, the European University Institute, Florence, Italy, the Dean of Studies and the Academic Service decided recently to introduce systematically the use of an anti-plagiarism software. The reason is for single Ph.D. researchers to look at the various chapters and drafts of their dissertation during the four years research/writing process and verify the originality of the contents. We want to avoid to have researchers be shamed and expelled out of the community of scholars like this student in Norway!

So, at the end of the process, when the thesis is submitted, each supervisor should perform this new task against plagiarism directly on the manuscript of his/her supervise. This task -and the instruments that are available to perform it- are today an evidence of the worldwide shift towards digital. It is taken for granted that everything we write is somewhere in the virtual space and can be retrieved and analyzed to avoid using someone else’s ideas without acknowledging it. This is an extraordinary shift in the humanities sciences towards “other” humanities. It introduced a bit of digital humanities for everybody in a way!

At the EUI, this task which was performed by the staff of the Dean of Studies and the Academic Service, has now to be performed directly by the thesis supervisor before the decision taken by the departments to officially accept that a candidate submit a thesis for discussion with the jury. The software Turnitin has been chosen and new administrative rules introduced on how to use it. Now, scholars on both side of the Ph.D. writing process, he who writes it and he who is supervising it, are both involved with digital tools. This is something that never happened before.

Introductory courses to plagiarism, originality check, good academic practices and, finally, to Turnitin itself, have
Using Turnitin beyond Plagiarism: THA TCamp Leadership 2013

Could we go beyond Turnitin & anti-plagiarism softwares?

The 10th of October 2013, I will participate to THATCamp Leadership at the RRCHNM, the Center for History and New Media at George Mason University, Fairfax, Virginia. THATCamp, The Humanities and Technology Camp was conceived at George Mason in 2008 and became soon international.

THA TCamps were held in Paris in 2010 and 2013, in Florence at the European University Institute in 2011, in Lausanne and in Luxembourg/Trier in 2012 and many other THATCamps in Europe and in other continents.

23rd - 26th of March 2011 Florence

Participating to THATCamp will allow you to perform Digital Humanities activities in informal ways. That’s why it has been called an unconference. Following the CHINMA’s definition, “an unconference is a highly informal conference. Two differences are particularly notable. First, at an unconference, the program isn’t set beforehand: it’s created on the first day with the help of all the participants rather than beforehand by a program committee. Second, at an unconference, there are no presentations — all participants in an unconference are expected to talk and work with fellow participants in every session.” During THATCamps “humanists and technologists of all skill levels learn and build together in sessions proposed on the spot”.

What should I propose to THATCamp Leadership is something I was wondering from sometimes now so I decided to post a session proposal on the GMU website looking at what are my next duties for the History Department at the European University
1. What is Turnitin?
2. What is plagiarism?
3. Your Turnitin account at the EUI
4. What is meant by Text Mining in Big Textual Databases?
5. How to submit a paper and interpret similarities with other texts?
6. Some examples
1. What is Turnitin?
What is Turnitin?

- Turnitin was conceived now 21 years ago in 1997 by John Barrie and three other students at the UC Berkeley to offer new forms of collective peerage and as an instrument to discover plagiarism in academic research.

- “The Turnitin software checks for potentially unoriginal content by comparing submitted papers to several databases using a proprietary algorithm. It scans its own databases, and also has licensing agreements with large academic proprietary databases”. (en.wikipedia)

- With Turnitin, you are reminded not to copy others without applying citation’s rules so to use the best way to quote a source in your papers. This is done looking at other papers quoting the same source through a «similarity search».

- In 2014, 26 million students and instructors used Turnitin globally and Turnitin's database reached 500 million submissions; more then 24 bilion web pages were indexed; more then 300 milion student papers archived in repositories worldwide were checked; more then 120 milion essays in books and journals were analyzed.
2. What is plagiarism?
What is a misconduct & plagiarism in academic research at the EUI?

**Misconduct** in academic research «implies (and is not limited to) fabrication, falsification, plagiarism or deception in proposing, carrying out or reporting results of research and deliberate, dangerous or negligent deviations from accepted practice in carrying out research.»

**Plagiarism** is the «deliberate copying of ideas, text, data or other work (or any combination thereof) without due permission and acknowledgement.»

*European University Institute: Code of ethics in academic research, v.2017 p.13*
We are guilty until proved innocent he said and refused to submit his papers to Turnitin.

The student also objected to the California-based company profiting from its database of student work.

Students against looking at plagiarism already in 2004!
available [here](#).

1. **Clone**: Submitting another's work, word-for-word, as one's own
2. **CTRL+C**: Contains significant portions of text from a single source without alterations
3. **Find – Replace**: Changing key words and phrases but retaining the essential content of the source
4. **Remix**: Paraphrases from multiple sources, made to fit together
5. **Recycle**: Borrows generously from the writer's previous work without citation
6. **Hybrid**: Combines perfectly cited sources with copied passages without citation
7. **Mashup**: Mixes copied material from multiple sources
8. **404 Error**: Includes citations to non-existent or inaccurate information about sources
9. **Aggregator**: Includes proper citation to sources but the paper contains almost no original work
10. **Re-Tweet**: Includes proper citation, but relies too closely on the text's original wording and/or structure
#9 Aggregator

Includes proper citation to sources but the paper contains almost no original work

**SOURCE TEXT**

A Natural Setting: A History of Exploration and Settlement in Yosemite Valley

Since its first discovery by non-indigenous people in the mid-nineteenth century, Yosemite Valley has held a special, even religious, hold on the American conscience because its beauty makes it an incomparable valley and one of the grandest of all special temples of Nature. While Yosemite holds a special grip on the western mind, perceptions about the Valley have evolved over time due to changing politics, migration patterns and environmental concerns as man has become more attuned to his relationship and impact on nature.

**STUDENT WORK**

A Natural Setting: A History of Exploration and Settlement in Yosemite Valley

Yosemite Valley was first sighted by non-Indians in 1833 by Joseph Rutherford Walker and his group of explorers. Yosemite has a long and diverse cultural history that helped to shape the region, country, and even the world. While Yosemite holds a special grip on the western mind, perceptions about the Valley have evolved over time due to changing politics, migration patterns and environmental concerns as man has become more attuned to their relationship and impact on nature.


4 Harrick, C. "A Natural Setting," p. 41.
How can you avoid plagiarism?

Rule 1:
Acknowledge direct use of someone else’s words.

Rule 2:
Acknowledge any paraphrase of someone else’s words.

Rule 3:
Acknowledge direct use of someone else’s idea(s).
Turnitin: checking for originality

What will **not** be identified as plagiarised

- phrases of 3 words or less
- ‘quotations’
- Bibliography

What **will** be identified as plagiarised

- block quotations without quotation marks
- identical passages
- partially paraphrased passages
3. What is meant by Text Mining in Big Textual Databases?
“Memory is not an instrument for surveying the past but its theater. It is the medium of past experience, just as the earth is the medium in which dead cities lie buried. He who seeks to approach his own buried past must conduct himself like a man digging.”

— Walter Benjamin, Berlin Childhood around 1900
Google Results
Google Scholars Results not using quotations marks: memory is not....
Google Scholars Results using quotations marks «memory is not....»
Google Books Results using quotations marks «memory is not....»
Google NGRAM Viewer: «Public History» in English books 1800-2008
Text mining in ECCO & EEBO databases XV-XIX centuries texts
Searching for «memory» in title in ECCO & EEBO
Text-Mining?

Open Access Example: Rousseau online
4. Your Turnitin account at the EUI
Originality Check - Turnitin Software

In support of the policy to promote good practice in academic research, EUI has purchased a licence for Turnitin - a plagiarism detector software. Plagiarism as defined in the EUI-Code of Ethics in Academic Research is "The deliberate copying of ideas, text, data or other work (or any combination thereof) without due permission and acknowledgement." (See also definitions of Research Misconduct pages 6-8 of the Code of Ethics in Academic Research).

As recommended by the EUI-Research Computing Committee and the Dean of Graduate Studies, EUI doctoral researchers and academics are invited to use the software not only as an anti-plagiarism software, but also as an educational instrument that can be used individually and within research networks to understand from where a quotation is coming, who already did cite/quote a phrase or a paragraph, etc. It is useful also if you, as an author, have already published a text you are re-using to understand where this happened.

The originality check feature of Turnitin has been activated for the EUI. This feature helps individuals submitting their own work to check for improper citation or potential plagiarism by matching it up with a wide text comparison database. Originality Check also reports on how much of the submitted text matches content from Turnitin databases so to indicate how much of the paper is original or unoriginal.

We will avoid the student/instructor interface model by allowing auto-submission of text, where once submitted, only you will be able to view the auto-generated originality report and your submitted paper. The originality report states where text in your submitted work matches other texts and provides links to the original sources.

The software does not submit your work to any repository. Your work will be compared with a quite wide range of online content and with some published texts, papers and books, but it does not cover all books and periodicals.

You can find more information about how to use Turnitin in the following link:
http://turnitin.com/en_us/training/getting-started
For any issues regarding Turnitin, contact the EUI Helpdesk at http://helpdesk.eui.eu

If you have any questions on Plagiarism, contact Nicola Owtram, Academic Service at nicola.owtram@eui.eu

Technical assistance is available at ICT service: Jot van der Meijden, Coordinator, Databases and Applications at jot.vandermeijden@eui.eu and Gabrielle Horan, Databases and Applications Software Engineer at gabrielle.horan@eui.eu
Dear All,

As a member of the EUI academic community, you will soon be receiving a standard message from Turnitin informing you that you have been enrolled in the ‘eUI originality check 2017–2018’ by the EUI.

Please note that you are not obliged to submit any of your work to the software. Your enrolment means that you will be able (if you so choose) to activate your account to submit your own work to the software. The originality reports will then be accessible only to you.

The originality check feature of Turnitin has been activated for the EUI. This feature helps individuals submitting their own work to check for improper citation or potential plagiarism by matching it up with Turnitin’s text comparison database. Originality Check also reports on how much of the submitted text matches content from Turnitin databases.

We will avoid the student/instructor interface model by allowing auto-submission of text, where once submitted, only the author will be able to view the auto-generated originality report and the submitted paper. The originality report states where text in your submitted work matches other texts in the Turnitin database and provides links to the sources.

The Turnitin software does not submit the papers further to any repository. A submitted paper will be compared with a fairly wide range of online content and with some published texts, papers and books, but it does not cover all books and periodicals and cannot detect plagiarism from sources not included in the Turnitin database, or the plagiarising of ideas.

More information about the EUI policies on Ethics and Integrity in Academic Research can be found on the following link: https://www.eui.eu/ServicesAndAdmin/DeanOfStudies/Ethics-and-Integrity-in-Academic-Research

Best wishes
Martin Scheinin
Dean of Graduate Studies
Gentile Noiret Serge,

Sei stato iscritto alla classe Turnitin class "EUI Originality check 2017 -2018" dal tuo docente, EUI Turnitin.

Turnitin OriginalityCheck consente ai docenti di verificare le citazioni improprie o il plagio potenziale, confrontando i lavori degli studenti con gli elaborati dei database in continuo aggiornamento. Ogni Originality Report fornisce ai docenti la possibilità di insegnare ai propri studenti i metodi di citazione appropriati, e di salvaguardare la loro integrità accademica.

Per iniziare questa classe, vai a http://www.turnitin.com/it/home, accedi usando il tuo indirizzo email e la password, e clicca sul link "EUI Originality check 2017 -2018" nella tua home page.

Ti ringraziamo per aver utilizzato Turnitin,

Il team di Turnitin
Guides to Turnitin

How can we help?

Manuals & Guides
Integrations
Release Notes
User Guides clicking on «help» in home page
Video Tutorials: how to use Turnitin?

About OriginalityCheck

Instructor Training

A typical submission made to an assignment on Turnitin generates an Originality Report. The Originality Report is the report that is used to check whether or not a submission is original. It contains information such as the percentage of similarity to academic sources, as well as suggestions for how to improve the submission.

Viewing Originality report

Instructor Training

The Originality Report can be viewed in one of four modes. These modes allow users to view and sort the information contained in the Originality Report in any way best suited to their needs. The four modes for viewing an Originality Rep
5. How to submit a paper and interpret similarities with other texts?
Logging in Turnitin:

http://www.turnitin.com/

Be careful not to use http://turnitin.com/

which is the UK branche of Turnitin
Who’s who in Turnitin at the EUI?

- «Instructors» = EUI Professors or EUI research group coordinators
- «Students» = everybody using Turnitin at EUI
- «Join an Account» = asking EUI Helpdesk if you are a fellow; researchers receive one at the start of each academic year
- Submit queries to EUI helpdesk
Benvenuto in Turnitin

Ciao Noiret Serge,

sei stato aggiunto come assistente alla classe EUI Library da EUI Turnitin.

Sei pronto a iniziare?

Fai clic di seguito per accedere e gestire questa classe.

In caso di problemi, visita il sito Web guides turnitin.com per consultare alcune linee guida utili.

Grazie,
Turnitin Home Page for EUI users at www.turnitin.com
Submit your paper online: [http://www.turnitin.com](http://www.turnitin.com)

- First originality report generated relatively quickly, depending on the size of the paper
- One paper at a time if student, many if instructors!
- As student previous submissions will be overwritten and they cannot be retrieved
- Originality reports for subsequent papers require a further 24 hours to be generated
- Texts are not stored in repositories or databases
- Only you can see your submission and also in text mode to access the sources when you are out of Turnitin
Accepted types of files?

- Microsoft Word,
- PowerPoint,
- WordPerfect,
- PostScript,
- PDF,
- HTML,
- RTF,
- OpenOffice (ODT),
- Hangul (HWP),
- Google Docs (submitted via the Google Drive submission option),
- Plain text files TXT
• **Turnitin** prepares an "originality report" on how the submitted work compares with other documents.
• It can also evaluate students' papers for spelling, grammar and structural errors.
• Could we use the software for text-mining and context searching?
What can we upload in Turnitin? Instructor’s page

Submit Paper: Multiple File Upload

Requirements for multiple file upload:
- Each file must be less than 40 MB (read suggestions to meet requirements)
- The maximum paper length for each file is 400 pages
- All file types are allowed and can receive marks in GradeMark and PeerMark; however, only Microsoft Word, PowerPoint, WordPerfect, PostScript, PDF, HTML, RTF, OpenOffice (ODT), Hangul (HWP), Google Docs, and plain text files with at least 20 words of text will be eligible to receive Originality Reports

File Submission Preview (editing fields is not required)

<table>
<thead>
<tr>
<th>Submit</th>
<th>File</th>
<th>Accepted?</th>
<th>Student</th>
<th>Last name</th>
<th>First name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Choose File No file chosen

Upload All
When a paper is submitted to Turnitin, it is compared against 3 content databases:

- **The Current and Archived Web**: Similar to Google and Bing, Turnitin has built a web crawler that crawls the Internet and indexes content into a searchable form. Turnitin currently contains over 24 billion web pages from the current web as well as archived web pages.

- **Student Papers**: Over 50 percent of plagiarism comes from other student’s work. Turnitin compares submitted papers to a database of over 300 million papers in the Turnitin paper database. Each day, the Turnitin student database grows by 190,000 papers.

- **Content Partnerships**: Turnitin has partnered with leading content publishers, including library databases, text-book publishers, digital reference collections, subscription-based publications, homework helper sites and books. These partnerships have contributed over 110 million additional articles to the databases.
After Submission of a Document in Turnitin
Dear Noiret Serge,

You have successfully submitted the paper, "test_for_HEC_course_8_10_2013" to the assignment "Self-check submissions" in the class "EUI originality check 2013 - 2014" on 04-Oct-2013 02:14PM. Your paper id is 358618486. Your full digital receipt can be downloaded from the download button in your class assignment list in Turnitin or from the print/download button in the document viewer.

Thank you for using Turnitin,

The Turnitin Team
Digital History 2.0

SERGE NOIRET

European University Institute, Florence, Italy

Introduction: a few Web 2.0 principles

Tim O’Reilly, who is generally credited with having spawned the term 'Web 2.0', said:

[...] one of the central differences between the PC era and the Web 2.0 era is that once the internet becomes platform [...] you can build applications that harness network effects, so that they become better the more people use them. I’ve used the phrase ‘harnessing collective intelligence’ to frame this phenomenon.¹
Digital History 2.0

SERGE NOIRET

European University Institute, Florence, Italy

Introduction: a few Web 2.0 principles

Tim O’Reilly, who is generally credited with having spawned the term ‘Web 2.0’, said:

‘...one of the central differences between the 1.0 era and the Web 2.0 era is that once the internet becomes platform [...], you can build applications that harness network effects, so that they become better the more people use them. I’ve used the phrase: harnessing collective intelligence. To frame this phenomena’.

With this point before us, we need to give thought to the changes that have come with the advent of new practices sprang from the use of second-generation Web technologies in the humanities and, where we are concerned, in history.
Filtering the originality report

- Click the Icon here
- You may exclude the bibliography and/or the quotes
- You may decide to change the default EUI settings to more than 4 words
- Abandon or not the multi-color highlighting
Originality Report Overview

- Shows percentage of sources matched
- Colour coded text to indicate matches to sources in database
- Compare your text with the source text
- Parameter filtering option (word matches, quoted text, bibliography)
- Print or download and save the report to your computer
Basic instructions: viewing your similarity report

Viewing the Similarity Report

About the Similarity Report

A typical submission made to an assignment in Turnitin generates a Similarity Report. The Similarity Report is the result of comparison between the text of the submission against the search targets selected for the assignment; this may include billions of pages of active and archived internet information, a repository of works previously submitted to Turnitin, and a repository of tens of thousands of periodicals, journals, and publications. Any matching or highly similar text discovered is detailed in the Similarity Report that is available in the Assignment Inbox.

The similarity score is a percentage of a paper’s content that matches to Turnitin’s databases; it is not an assessment of whether the paper includes plagiarized material.

Viewing similarity matches

Table of contents

1. About the Similarity Report
2. Viewing similarity matches
3. Viewing sources
4. Filtering the Similarity Report

4.1. Excluding quotes and bibliography
4.2. Excluding small matches
4.3. Similarity Report multi-color highlighting

5. Generating a new Similarity Report
6. Excluding Sources

6.1. Excluding sources via match overview
6.2. Excluding sources via all sources
Basic instructions: interpreting your report

Interpreting the Similarity Report

Turnitin does not check for plagiarism in a piece of work. Instead, we will check a student's work against our database, and if there are instances where a student's writing is similar to, or matches against, one of our sources, we will flag this for you to review. Our database includes billions of web pages: both current and archived content from the internet, a repository of works students have submitted to Turnitin in the past, and a collection of documents, which comprises thousands of periodicals, journals, and publications.

Table of contents
1. Similarity Score
2. Refining a Similarity Report
3. Similarity Scoring Examples

Match Overview
89%

It is perfectly natural for an assignment to match against some of our database. If your student has used quotes and has referenced correctly, there will be instances where we will find a match. The similarity score simply makes you aware of any problem areas in a student's paper; you can then use this as a tool as part of a larger process, in order to determine if the match is or is not acceptable.

Similarity Score
6. Some examples
Using Turnitin to Contextualize Citations

4 Tests: 3 essays, one quotation
Confronting my text with other sources

- Click the number
- The publication opens in a pop-up window
- You may ask to view the Full Source
Viewing the full paper after originality check isn’t always possible

- This would be the same for EUI papers loaded in the database
- The EUI Dean of Study and EUI Academic Service policy is to avoid our papers becoming part of the Turnitin database
“Memory is not an instrument for surveying the past but its theater. It is the medium of past experience, just as the earth is the medium in which dead cities lie buried. He who seeks to approach his own buried past must conduct himself like a man digging.”

— Walter Benjamin, Berlin Childhood around 1900
Review of *History Manifesto* (4 authors) for *Memoria e Ricerca*. Today a portion of the text has been republished in a blog and the 4 authors can trace it as "The History Manifesto: a discussion", introduction by Serge Noiret, with contributions by Ramses Delafontaine (editor), Quentin Verreycken, Eric Arnesen, in "Memoria e Ricerca", 1/2016, pp. 97-126.
Clicking on source n.24 to discover essays using same contents

Clicking on source n.25 to discover essays using same contents
Clicking on n.9 «eac.revues.org» to view similarity in the source, an academic journal in revues.org
Source n.9 may potentially extend my research area with new literature on the «Comité de Vigilance...»
Cristina BLANCO SÍO-LÓPEZ: *Inner Tensions vs. Outer Change: Political Agency Impact on Cohesion Funds Implementation in Spain in the Context of the Maastricht Treaty* in *Journal of European Integration History* 2, 2017

Test 4: Accessing the whole text quoted in the paper in the source in turnitin (right column) or outside turnitin
Do you have any further question?

Thank you!