Software support for originality checks

What is plagiarism?

- Plagiarism is using someone else's work or ideas without attributing proper credit and presenting the work or ideas as your own

- The word "to plagiarize" is Latin in origin and means "to kidnap."

- Extreme cases:

  - Minor case: the quotation of a sentence or two, without quotation marks and without a citation to the true author

  - Major case: a significant fraction of the entire work was written by someone else e.g. "the plagiarist removes the true author names and substitutes the plagiarist's name, then submits the work for credit (e.g., term paper or essay) or as part of the requirements for a degree (e.g., thesis or dissertation), or even as a professional/personal concern (e.g., article, blog entry, news item, etc.)" [ckeckforplagiarim.net]

Types of plagiarism

Type of plagiarism	Definition	Severity
Global or complete plagiarism	Presenting an entire text by someone else as your own work.	Severe
Outsourcing	Paying to write on behalf	Severe
Paraphrasing plagiarism	Rephrasing someone else's ideas without citation.	Serious
Verbatim or direct plagiarism	Directly copying a passage of text without citation.	Serious
Mosaic plagiarism	Combining text and ideas from different sources without citation.	Serious
Inaccurate or misleading authorship	An individual contributes to a document but does not get credit.	Serious
Source-based plagiarism	Reference an incorrect or non-existing source	Serious
Self or auto plagiarism, duplication	Reusing passages and ideas from own previously submitted work.	Moderate
Incorrect citation	Failing to give all the necessary information in the source citation.	Moderate
Accidental plagiarism	Occurs because of neglect, mistake, or unintentional paraphrasing	Moderate

Other close related problems

- Patch writing: when too many quotations are included too close together without providing any context around them.

- Secondary source: happens when it is cited a document that only cites a primary source.

- The previous table & others cases can be reformulated (not in the same order):

  "clone", "remix", "Ctrl+C", "hybrid", "find-replace", "bluffing",

  "recycle", "mashup", "404 error", "aggregator", "re-tweet", "stitching"
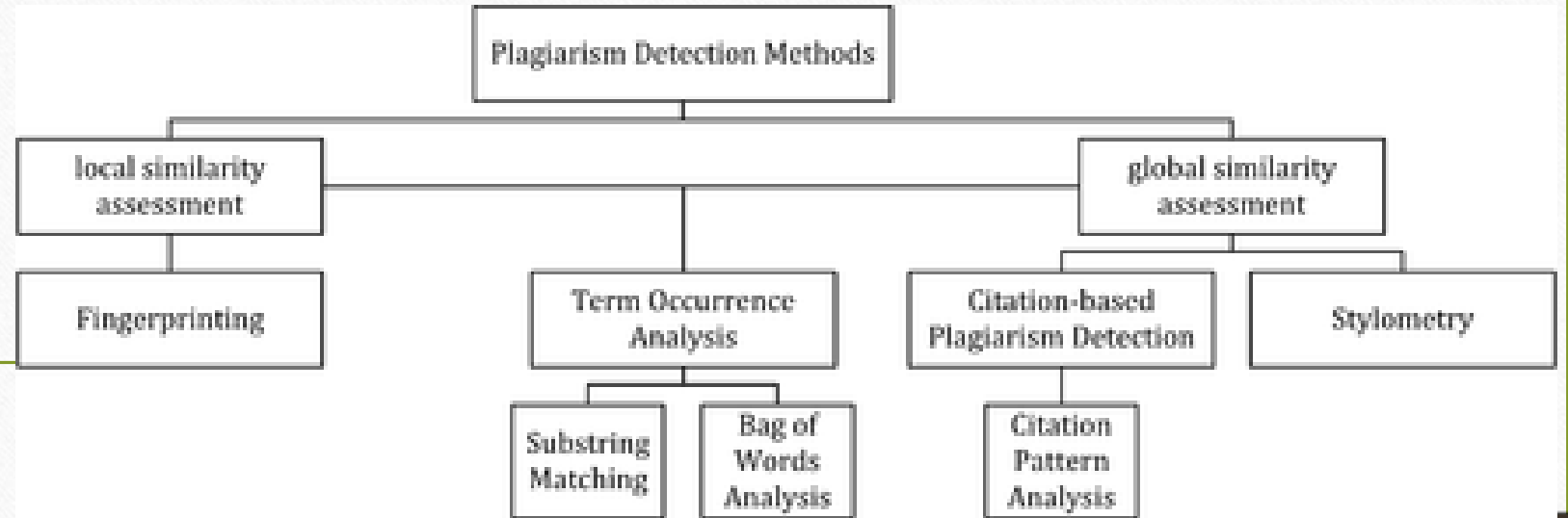
Why?

- Plagiarism of even a sentence or two can have serious consequences.

- Legally is considered a fraud.

- In an academic context, it is a very serious ethical offense.

- For students, plagiarism often means a failing grade, academic probation, or worse.

- For researchers, is a shame that can lead to a change in career
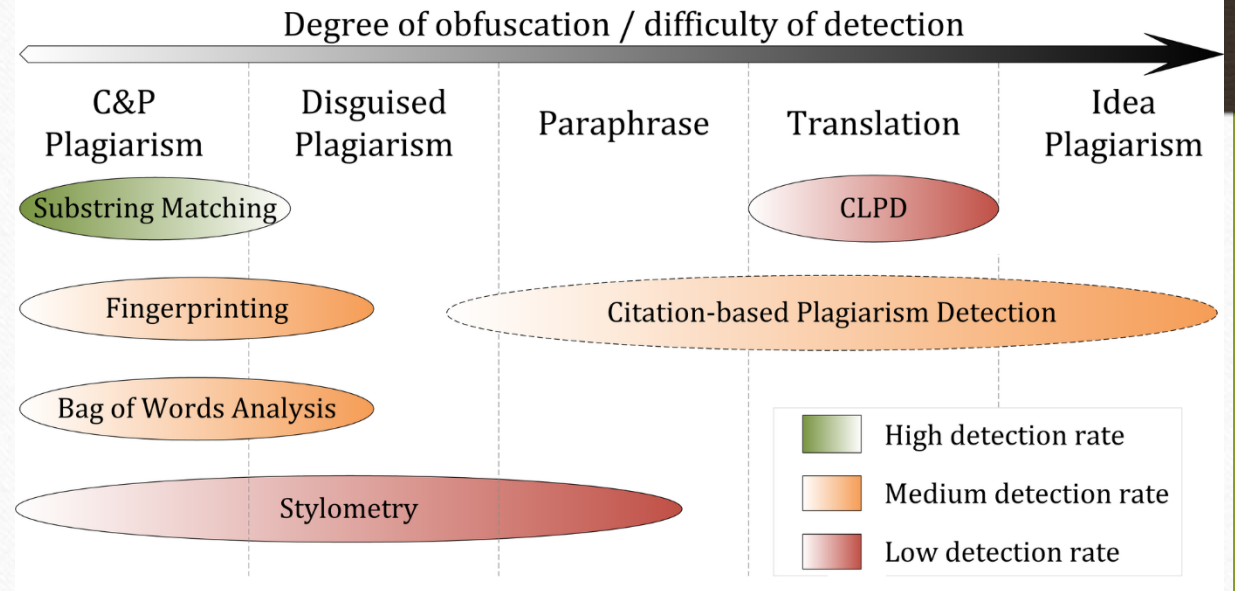
Avoiding plagiarism

- Keep track of the **sources** that are consulted in the research.
- **Paraphrase or quote** from the sources, but add own ideas.
- Credit the original author in an in-text **citation** and reference list.
- Use a **plagiarism checker** before submit.

Plagiarism detection methods

- Fingerprint: using digests of documents
- String matching: using suffix document models
- Bag of words: vector space retrieval
- Citation analysis: similar patterns in citation sequences
- Stylometry: statistical models

Anti-plagiarism software - criterias

Factor	Description and alternatives
Scope of search	In the public internet, using search engines / institutional databases / local, system-specific database.
Analysis time	Delay between the time a document is submitted and the time when results are made available
Capacity	No. documents the system can process per unit of time
Check intensity	How often and for which types of document fragments (paragraphs, sentences, fixed-length word sequences) does the system query external resources, such as search engines.
Algorithm type	The algorithms that define the way the system uses to compare documents against each other.
Precision and Recall	No. documents correctly flagged as plagiarized compared to the total no. flagged documents, and to the total no. documents that were actually plagiarized.

Commercial tools

- iThenticate (& CrossCheck)

- Turnitin

- PlagAware

- PlagiarismDetection

- PlagiarismSearch

- Scribbr

- Grammarly

- Viper

- DrillBit

…

Three examples

Turnitin

- Gives the most accurate similarity index at present, including math texts
- Automatically stores uploaded texts (submitted for checking) in its own database

Scribbr

- developed specifically for students
- largest plagiarism database
- the checked document can be deleted
- high accuracy
- in partnership with Turnitin

Grammarly

- integrated into the speller and grammar checker
- easy to use
- Submissions are limited to 1000 words in freenium version
- Checks againsts web pages and ProQuest database
- low accuracy

Free plagiarism checkers

- Copyleaks : 2500 words/months for free

- PaperRater: 10 checks per month

- Plagium: max 1000 characters/checks

- StudyMoose: small documents

- Oxphrase: 30000 works/check, credit system

- Unicheck (former Unplag): 500 words/free

- Duplichecker: 50 scans per day

- SmallSEOTools: max. 1000 words/check

- SearchEngineReports: max. 1000 w/check

- Prestoseo: max 1000 words/check

- Edubirdie: free (but can sell the text)

Trial free

- Quetext: free trial 3 x 500 words

- EasyBib: free trial 3 days

- PlagScan: free trial 2000 words

- Plagramme: free procent, pay for sources

- Copyscape: 10 free tries

- Plagiarism Checker X: free for 14 days

- Plagiarisma: limited number of free checks

- CheckForPlagiarims: free trial for students

- PlagTracker: upload file not available in free version

- WhiteSmoke: free trial for 4 days

Open source

For text

- Sherlock
- Plagiarism-checker
- Whizzygeeks for WordPress
- WCopyfind compare owned documents

For codes

- OSPC
- SIM
- JPlag