

UCAS Similarity Detection Service - guidance for applicants

This guide is designed to help UCAS applicants understand our similarity detection process. All personal statements sent to UCAS are tested for similarity.

There are some example personal statements on the internet that have been used by applicants, in some cases word for word. The service we use, called Copycatch, finds statements that show similarity, works out how much of the statement may have been copied, and reports the findings. It helps admissions staff at universities and colleges judge applications, and it is the institutions who decide what action, if any, to take regarding notified cases.

Research has shown that the majority of UCAS applicants do write their own personal statements. However, the number making use of other people's material was high enough to justify the introduction of the Similarity Detection Service.

What the Similarity Detection Service does

Each personal statement is checked against:

- a library of personal statements previously submitted to UCAS
- sample statements collected from a variety of websites
- other sources including paper publications

Each personal statement received at UCAS is added to the library of statements after it has been processed.

What happens if a personal statement has similarities?

- Any statements showing a level of similarity of 10% or more are reviewed by members of the UCAS Similarity Detection Service Team.
- Applicants, universities and colleges are notified at the same time by email when an application has similarities confirmed.
- Admissions tutors at individual universities and colleges decide what action, if any, to take regarding reported cases.

Eliminated words

The Copycatch process ignores commonly used words that many applicants use in their statements such as 'and', 'so' and 'with'.

Copycatch also ignores a selection of commonly used words and phrases including 'Duke of Edinburgh' and 'football'.

The verification process

- Copycatch identifies sentences in a personal statement that are matched to other personal statements already held in the Copycatch system.
- Levels of similarity are checked by trained staff who decide whether you and the institutions you are applying to need to be informed that similarity has been found.
- The universities and colleges you are applying to decide on the significance of the results and what action, if any, to take.
- Your personal statement will not be compared to your earlier applications, if you have applied in previous cycles or schemes.

Notification that a report has been sent to the universities and colleges

If Copycatch finds a significant level of similarity in your personal statement and the Verification staff at UCAS decide to inform the institutions you have applied to, we will let you know by email (if you have a verified email address). This email includes instructions on how you can view what Copycatch has found in Track, and gives you a link to frequently asked questions for further advice and guidance.

The report sent to you is identical to the report sent to the institutions. It displays your personal statement marked up to identify sentences similar to others in the Copycatch system.

How we show matches with other statements

We use four colours (see below) to indicate significant matches with other statements and grey to show sentences which have not been found to match.

Within matched sentences, words which are different from the one matched with it by the program are highlighted in **black**. **Underlined black** is used to show that the word is related but not identical.

What the sentence colours mean

Red is used for the sentences from the most matched statement.

Blue is used for the next best match if there are at least three sentences.

Pink is used for the third best match if there are at least another three sentences.

Brown is used for any other matches if there are at least three sentences.

Grey is used for sentences for which no match has been found and for very short sentences, which don't get checked.

Examples

I grew up in a city near the sea and have always been fascinated by marine life.

If you had written this sentence and found it shown in red as above when you checked the notification report, it would mean that it had been exactly matched to a personal statement stored in the Copycatch library.

I grew up in a town near the sea and have always found marine life fascinating.

If the sentence you had written was marked in your report like the one above, it would mean that:

- **town** and **found** were not in the matched sentence
- fascinating was not found as an exact match but is similar enough to the equivalent word in the matched sentence to be identified by underlining

The blue colour also shows you that the match was found in the second most matched statement.

The dates on the matched personal statements

At the bottom of the marked up personal statement, the number of sentences matched to library or internet sources is shown in the same colour as that used to mark up the sentences.

The date shows how long this personal statement has been in the UCAS collection. It does not mean that this particular statement was the one used as the source for the current personal statement.

Both may be taken from a source outside the library, or there may be other related files inside the library which have not been shown because there was no additional matched information.

The dates on the matching web sources

The number of web source sentences is shown in the same way, but here the date means either the date it was posted to the website, if known, or the date when the web source was identified by UCAS. Again, it does not necessarily mean that the file was the actual source.

As a feasibility study discovered, some web sources are very popular, and may appear on more than one website, or have been used in a modified form in a personal statement within the UCAS collection.

Why the program works

- A personal statement of 4,000 characters will contain approximately 600 words, about half of which will be words that are eliminated from consideration (see above).
- Usually, if two personal statements are randomly selected and compared, you would expect very little or no similarity. Most sentences will be significantly different.
- This means that if Copycatch finds two sentences in different statements which have exactly the same words, it is very likely that one is a copy of the other, or that both have been copied from a third source. Of course this can and does happen in essays if a quote from a text is included, but is very unlikely to occur in a personal statement.
- If Copycatch finds a number of identical or similar sentences in a personal statement and a file held in the library, then a similarity report is generated.