Anti-counterfeiting and anti-piracy technology guide

https://euipo.europa.eu/anti-counterfeiting-and-anti-piracy-technology-guide

Hashing

Hashing is a process that uses an algorithm (such as MD5, SHA-1 or SHA-2) to create a unique identifier – a 'hash' – for a file on the basis of its data. Two identical files will always have the same hash. Likewise, two different files will always have different hashes, even if the difference between the two files is minimal. The hashes of files found to infringe copyrighted material can be added to 'black lists', which can be used by cloud storage services to detect, block and take down unauthorised files.

Used for

- Authentication

Goes on

- Product

Can be used for

- Digital products

Visibility

- Invisible

Reading device needed

- Yes

Connection to server

- Yes
- No

Description

The purpose of hashing is to identify different copies of the same digital file. First, an algorithm is used to assign a unique string of characters, called a 'hash', to the digital file, and that hash is stored in a hash database. A second digital file is then put through the same hashing process, and the resulting hash is compared with the one already stored in the hash database to determine whether the files match exactly.

Uses

Hashing is primarily used to index content, secure passwords, detect malicious files such as viruses, or block illegal content, such as the promotion of acts of terrorism or exploitation of minors. With respect to copyrighted material, hashes serve to identify, compare and prevent the upload of digital content that has already been taken down due to copyright infringement.

Implementation

It is difficult to provide specific implementation requirements for hashing as there are several different hashing techniques and algorithms that can be used, all of which require different amounts of computational power. Most hashing methods are cryptographic (in which two different files, although very similar, will have completely different hashes) but perceptual hashing methods also exist (in which two almost identical files, for instance similar images, will have similar hashes). That said, the advantage of hashing is that it can be applied to already existing content. Hashes can be shared and used by all the different parties using the same hashing process and the same hash database. The disadvantage, however, is that hashing only recognises the same identical file, not its content.

Cost

Hashing requires limited investment as open source solutions are available. Furthermore, as hash databases only store a string of characters rather than the entire file, they require fewer computational and memory resources.

*The information contained in this site is of **general** informational nature and does not constitute legal, consulting or any other **professional advice or guidance**. The information contained in this site is provided on an 'as is' basis with **no guarantee of completeness**, accuracy, usefulness or timeliness and without any **warranties** of any kind whatsoever, express or implied, including but not limited to warranties of performance, merchantability and fitness for a particular purpose. The use of the information is at the user's own risk and the EUIPO does **not guarantee any specific results** or outcomes based on the use of the information provided on this website. The search results are **not presented in any order of recommendation** . This website contains links to **other third-party websites**. Such links, which are free and publicly available, are only for the convenience of the user or browser; the EUIPO does **not recommend, endorse**, warrant, or assume liability of the content or practices of these third-party sites. The EUIPO may update, modify, or remove any information on this website without prior notice. It is the user's responsibility to stay informed of any changes.*