

11. Mechanizmy kontroli spójności systemu dLibra

System dLibra wykonuje okresowo pewne czynności, które ułatwiają kontrolę spójności danych przechowywanych w systemie. Jeśli system napotka na niespójność i nie może samodzielnie rozwiązać tego problemu to komunikuje ten fakt administratorowi systemu. W systemie dLibra wykonywane są następujące czynności kontrolujące spójność systemu:

- Kontrola spójności plików publikacji - system dLibra sprawdza okresowo, czy pliki publikacji zostały zmodyfikowane bez wiedzy systemu. W momencie wprowadzania plików do biblioteki cyfrowej system dLibra oblicza tzw. **sumy kontrolne** dla każdego z plików. Następnie okresowo (domyślnie co noc) system weryfikuje, czy pliki którymi zarządza mają nadal taką samą sumę kontrolną jak w momencie ich przesłania do systemu. Jeżeli plików w bibliotece cyfrowej jest niewiele, to możliwe jest, że każdej nocy system zweryfikuje zawartość wszystkich plików. Jeżeli plików jest na tyle dużo, że weryfikacja całości danych trwa dłużej niż liczba godzin weryfikacji przewidziana na jedną noc, system każdej nocy będzie sprawdzał te pliki, które były weryfikowane najdawniej. Operacja weryfikacji kończy się podsumowującym wpisem w logach serwera, np.:

```
INFO: Consistency check finished. Took 24423717 ms.  
Not all versions (only 780) were checked.  
24 of total 780 compared versions and stored checksums is probably damaged.
```

W tym przypadku widać, że operacja sprawdzania spójności zajęła dokładnie 24423717 ms, czyli około 6h i 47 minut. Nie udało się w tym czasie sprawdzić wszystkich plików, sprawdzono ich tylko 780. Wśród tych 780 plików 24 miały sumę kontrolną niezgodną z sumą wzorcową zapisaną w bazie danych.

Dodatkowo w czasie przeprowadzania kontroli spójności dla każdego pliku, którego aktualna suma kontrolna nie zgadza się z sumą wzorcową do logów serwera zapisywana jest dodatkowa wiadomość. Logowana jest również sytuacja, w której dostęp do pliku okazuje się niemożliwy np.:

```
ERROR: Problem while calculating digests  
java.io.FileNotFoundException:  
F:\dlibra\content\files\1.dir\5.dir\22.dir\696.grp\664.pub\76075.fil\76283.ver  
(The system cannot find the file specified)
```

Zawarta w wiadomości ścieżka wskazuje, o jaką publikację i o jaki dokładnie plik chodzi. W powyższym przykładzie mamy np. do czynienia z plikiem (.fil) o id 76075, a konkretnie jego wersją (.ver) o id 76283. Plik ten stanowi część publikacji (.pub) o id 664, która jest częścią publikacji grupowej (.grp) o id 696. Publikacja ta znajduje się w ścieżce trzech katalogów (.dir) o następujących id: 1, 5 i 22. Informacje te są wystarczające aby odnaleźć plik z poziomu Aplikacji Redaktora. Można również wykorzystać identyfikator publikacji do dostępu przez stronę WWW. W tym celu należy do głównego adresu biblioteki cyfrowej dodać `/publication/<id publikacji>`.

- Kontrola spójności indeksów wyszukiwawczych - system dLibra sprawdza, czy indeksy wyszukiwawcze są spójne z metadanymi indeksowanych obiektów. Jeśli nie ma takiej spójności w przypadku konkretnego obiektu to jest on poddany ponownej indeksacji (co naprawia niespójność).