

АНОТАЦІЯ

Інформаційні технології пошуку плагіату в електронних документах // Дипломна робота // Бугальський Дмитро Ігорович // Тернопільський національний технічний університет імені Івана Пулюя, Факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних систем та мереж, група СІм-61 // Тернопіль, 2014 // с. – 105, рис. – 14 , табл. –6 , аркушів А1 –8 , додат. –2 , бібліогр. – 54 .

Ключові слова: плагіат, токен, електронний документ, програмний код, шингл.

Робота присвячена дослідженню способів контролю присутності плагіату в електронних документах. Рішення основного завдання в першу чергу пов'язано з розробкою відповідного методологічного та алгоритмічного забезпечення. Об'єкт дослідження – інформаційні технології пошуку плагіату у електронних документах. Предмет дослідження – засоби, методи і алгоритми автоматизованого пошуку плагіату у електронних документах.

В роботі запропонований і реалізований у вигляді самостійного модуля алгоритм пошуку запозичених фрагментів в електронних документах, що можуть містити вихідні коди програм, інтегруючий структурний аналіз кодів (на основі токенів) і методи шинглів, дистанції Левенштейна і найбільшої загальної підпоследовності (longest common subsequence, LCS) для аналізу довільних текстів.

ABSTRACT

Bygalsky D. I. Information technologies of search of plagiarism in electronic documents.

Work for the degree of Master of Science specialization 8.05010201 - Computer networks and systems - Ternopil Ivan Pul'uj National Technical University, Department of Computer Information Systems and Software Engineering, Computer Networks and Systems, Ternopil, 2014.

Work is sanctified to research of methods of control of presence of plagiarism in electronic documents. The decision of basic task is first of all related to development of the corresponding methodological and algorithmic providing. A research object is information technologies of search of plagiarism in electronic documents. The article of research is tools, methods and automated search of plagiarism engines in electronic documents.

In-process the search of the adopted fragments engine offered and realized as an independent module is in electronic documents, which can contain initial коди of the programs, integrating structural analysis of kodas (on the basis of токенів) and methods of шинглів, distances of Levenshtein and most general subsequence (longest common subsequence, LCS) for the analysis of arbitrary texts.

Keywords: plagiarism, token, electronic document, program code, shingle.