Приложение Б
Обзор источников

Название работы	Авторы, год публикации	Методы, использованные в работе	Описание данных	Достигнутая точность классификации	Язык программирования
Using classification techniques to determine source code authorship	B. Pellin, 2008	ACT, SVM	4 схожие программы, 2 автора	67 — 88 %	Java
Source code authorship attribution using n-grams	S. Burrows, S. Tahaghoghi, 2007	N-граммы	Выборка из 1640 файлов исходного кода и 100 авторов	67 %	C
A Probabilistic Approach to Source Code Authorship Identification	J. Kothari, M. Shevertalov, E. Stehle, S. Mancoridis, 2007	Статистический подсчет символьных метрик, составление на их основе профиля автора, обучение классификатора (Байес, VFI)	1) 12 авторов, 2110 файлов из open-source проектов; 2) 8 студентов, 220 файлов	70 — 90 %	Не указан
Identifying Authorship by Byte-Level N-Grams: The Source Code Author Profile (SCAP) Method	G. Frantzeskou, E. Stamatatos, S. Gritzalis, 2007	Составление профиля программиста на основе статистических метрик, подсчет отклонения от профиля	Не указано	88 % для C++, 100 % для Java	Java, C++
Application of information retrieval techniques for source code authorship attribution	S. Burrows, A. Uitdenbogerd, T. Urpin, 2009	N-граммы, рейтинговые схемы	100 авторов, классифицировались по 10, 1579 программных файлов	77 %	C

Продолжение таблицы

Название работы	Авторы, год публикации	Методы, использованные в работе	Описание данных	Достигнутая точность классификации	Язык программирования
De-anonymizing Programmers via Code Stylometry	A. Caliskan-Islam, R. Harang, A. Liu, F. Yamaguchi, 2015	Статистический подсчет признаков, нечеткие АСТ	250 авторов, 1600 файлов	94 — 98 %	C/C++, Python
Git Blame Who?: Stylistic Authorship Attribution of Small, Incomplete Source Code Fragments	A. Caliskan-Islam, E. Dauber, R. Harang, R. Greenstadt, 2017	Калибровочные кривые, нечеткие АСТ, классификатор Random Forest	Некомпилируемые неполные образцы кода с ресурса GitHub	70 — 100 %	C/C++