

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2020/151173 A1

(43) 国际公布日
2020 年 7 月 30 日 (30.07.2020)

(51) 国际专利分类号:
G06F 21/64 (2013.01)

(21) 国际申请号: PCT/CN2019/092293

(22) 国际申请日: 2019 年 6 月 21 日 (21.06.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910075454.8 2019 年 1 月 25 日 (25.01.2019) CN

(71) 申请人: 深信服科技股份有限公司 (SANGFOR TECHNOLOGIES INC.) [CN/CN]; 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。

(72) 发明人: 杨荣海 (YANG, Ronghai); 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋

一层, Guangdong 518055 (CN)。王大伟 (WANG, Dawei); 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。黄志伟 (HUANG, Zhiwei); 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。何嘉伟 (HE, Jiawei); 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。

(74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) Title: WEBPAGE TAMPERING DETECTION METHOD AND RELATED APPARATUS

(54) 发明名称: 一种网页篡改检测方法及相关装置

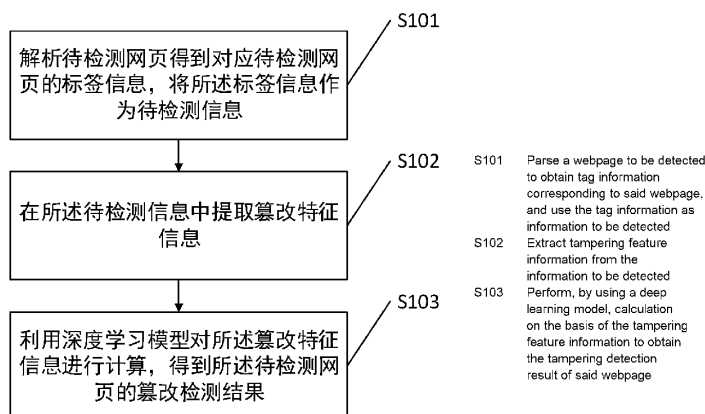


图 1

(57) Abstract: A webpage tampering detection method and a related apparatus. The method comprises: first parsing a webpage to be detected to obtain corresponding tag information; performing tampering feature extraction on the tag information to obtain tampering feature information; and performing, by using a deep learning model, calculation on the basis of the tampering feature information to obtain a tampering detection result. In the solution, a deep learning model and tampering feature information in a webpage to be detected are used for webpage tampering detection, and the deep learning model does not perform simple matching in a calculation process and has certain associative ability because the model is trained according to a training sample corresponding to tampering features. Therefore, if tampering feature information is not identical to the trained model, a more accurate detection result can be obtained, thereby improving webpage detection ability.

[见续页]

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种网页篡改检测方法及相关装置, 该方法包括先解析待检测网页得到对应的标签信息, 对标签信息进行篡改特征提取, 得到篡改特征信息, 并用深度学习模型对篡改特征信息进行计算, 得到篡改检测结果。由于本方案中利用深度学习模型和待检测网页中的篡改特征信息对该网页进行篡改检测, 而深度学习模型由于是根据篡改特征对应的训练样本训练出的模型, 在计算过程中不是简单的匹配, 而是具有一定的联想能力, 当篡改特征信息与训练模型不完全相同时也可以得到更准确的检测结果, 从而提高网页检测能力。