

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
14 February 2019 (14.02.2019)



(10) International Publication Number
WO 2019/032660 A1

- (51) International Patent Classification:
G06F 17/30 (2006.01)
- (21) International Application Number:
PCT/US2018/045740
- (22) International Filing Date:
08 August 2018 (08.08.2018)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/542,631 08 August 2017 (08.08.2017) US
15/837,518 11 December 2017 (11.12.2017) US
- (71) Applicant: AB INITIO TECHNOLOGY LLC [US/US];
201 Spring Street, Lexington, MA 02421 (US).
- (72) Inventor: FREUNDLICH, Robert; 55 Maple Ave., Sud-
bury, MA 01776 (US).
- (74) Agent: ATTISHA, Michael, J.; Wolf, Greenfield & Sacks,
P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).
- (81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, 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) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (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,

(54) Title: TECHNIQUES FOR DYNAMICALLY DEFINING A DATA RECORD FORMAT

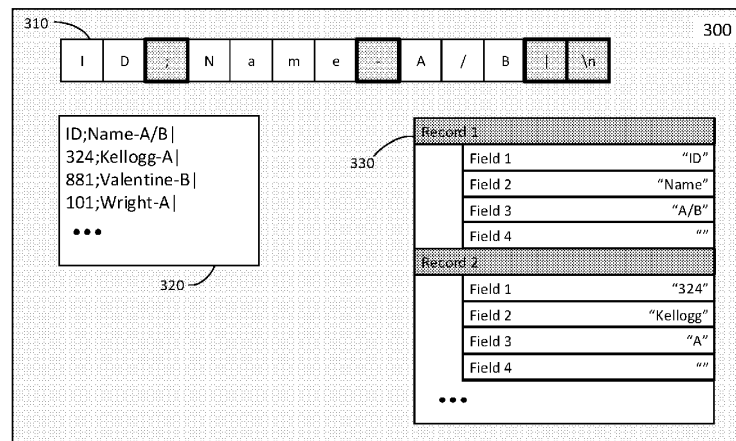


FIG. 3B

(57) **Abstract:** According to some aspects, a tool is provided that reduces errors made by a data processing system by assisting a user in determining a record format for a dataset by dynamically analyzing contents of the dataset based on real-time feedback provided by the user. The data processing system may apply the determined record format to automatically parse contents of the dataset, with fewer errors. According to some aspects, the tool may generate a user interface that allows a user to identify delimiters based on the content of the dataset, and may generate a provisional record format according to the identified delimiters.

[Continued on next page]



WO 2019/032660 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*