Multilingual Name Matching
Application to KYC


Marius Frunza
marius.frunza@schwarzthal.com


Ilgiz Mustafin
ilgiz.mustafin@schwarzthal.com


George Emelyanov
g.emelyanov@schwarzthal.com

Synopsis

Assessing the fully-fledged picture of a client is a crucial aspect of the KYC process. This assessment becomes complex when the client is resident and owns businesses in two or more countries. The compliance officer needs to find all the records for such clients in different databases and registries. When dealing with clients having data recorded in different languages and with different alphabets, the simple name matching methods have severe limitations.

For instance, a name like *Serge Pougacheff* can appear with very different forms when transliterated in Russian, English and French. Besides the original Cyrillic version Сергей Викторович Пугачёв, the Latin transcription includes: Sergei Pugachev, Serguei Pougacheff, Serge Pugachoff, Sergey Pugachyov, Serguei Pougatchev or Sergey Pugachew. Matching all possible forms of a name is the task of multilingual name matching, that is a driver in an efficient KYC process.

This paper analyzes a few optimal methods of multilingual entity matching used for entity resolution. The primary name matching approach using string comparison metrics is enriched with phonetic rules and with relational information. When applied in practice to solve KYC issues, most entity matching methods generate a significant amount of false positives. We enrich the current methods with a Bayesian approach based on the distribution of the frequency of name occurrence in a given language. The results applied to names of companies' directors from British and Russian business registries show that the approach using transliteration enhanced by phonetic matching and Bayesian search provides with the best performance

1. Introduction

Over the past five years, the banking sector has been hit by a wave of penalties for serious deficiencies in anti-money laundering and anti-terrorist financing systems and processes. These shortcomings are not solely the result of the propensity of some banks for customers or risky schemes. They also reflect the inefficiency of the compliance tools currently used by financial institutions, including KYC (Know Your Client). KYC is the process of a business identifying and verifying the identity of its clients.

Several banks (Table 1), including Danske Bank and Deutsche Bank, have suffered serious consequences, among other things because of shortcomings in their KYC methodology. Nevertheless, in many cases, the clients of these banks did not have a risky profile, but they had holdings in foreign companies, or they were associated with people subject to sanctions or involved in illegal activities. Therefore it is crucial in the KYC process to have a fully fledged picture of an individual, or a company with their corresponding transnational networks. The information concerning companies and directors are in many cases available on the national business registries. An advanced KYC should be able to access data from those sources and to map information about a person or entities from few different sources, containing data in different languages or different alphabets.

This process is called entity resolution also known as record linkage (Winkler (1999)), reference reconciliation (Saıs et al. (2007)) or object matching. It denotes the task of finding records from one or multiple databases, referring to the same real world entity Singla and Domingos (2006). Entity resolution in a single database case is sometimes called duplicate detection or deduplication Christen (2012).

As the amount of available data from national business registries increases entity resolution requires more resources and attention. In practice compliance manager in banks distinguish entities (person or companies) by themselves based on a manual process requiring human intervention. But for an efficient KYC process it is crucial to link entities across all business registries and data sources in order to provide compliance managers with advanced intelligence.

Name matching techniques are essential for joining data from different sources, especially from different business registries. Using exact string matching is not enough because one person can have multiple version of the names when transalted in different languages. As an example the name of the Russian ex-oligarch Sergei Viktorovich Pugachev appears in the Russian, English and French business registries in few versions. Beside the original Cyrillic version Сергей Викторович Пугачёв, the latin trascription include: Sergei Pugachev, Ser-

Bank/Country	Year	Penalty	Facts
Pilatus Bank (Malta)	2017	Liquidation	The bank has organized schemes to evade US sanctions against Iran.
HSBC (Hong Kong)	2018	0.35 Bil. USD	The bank has helped wealthy clients to avoid paying taxes..
HSBC (Hong Kong)	2012	1.9 Bil. USD	Cartels of drug traffickers laundered funds through bank branches in Latin America.
ABLV Bank (Latvia)	2018	Liquidation	The bank facilitated money laundering through illicit transactions for sanctioned entities in North Korea, Azerbaijan, Russia and Ukraine..
Deutsche Bank (Germany)	2010-2019	0.5 Bil. USD	The bank helped clients create offshore companies in tax havens to launder money.
Danske Bank (Danemark)	2007-2015	To be determined	The Estonian branch had transferred 235 billion USD largely to suspect customers in Russia and other former Soviet republics.
UBS (Switzerland)	2011-2013	0.02 Bil. USD	The bank's supervisory analysts would have cleared the alerts even as the transactions emitted warning signals.
US Bancorp (US)	2018	0.6 Bil. USD	The bank was investigating only a very limited number of suspicious transactions.
ING (Holland)	2010-2016	0.9 Bil. USD	The bank was convicted of non-prevention of money laundering and bribery, including bribery paid to the daughter of the Uzbek president by a unit of a Russian mobile phone operator.
BNP Paribas (France)	2004-2012	8.9 Bil. USD	The bank has put in place schemes that allow clients to circumvent sanctions against entities in Iran, Sudan and Cuba.

Table 1: List of the principal banks fines for KYC/AML inefficiencies.

guei Pougacheff, Serge Pugachoff, Sergey Pugachyov, Serguei Pougatchev. His son Alexander Pugachev appears also under Alexander Pugachew. Matching these possible forms of the name is the task of name matching, that is required in an efficient KYC process.

Applying name matching techniques has some difficulties. Usually, name similarity functions are designed to measure similarity between two words (two first names, two last names, etc.) and not between full names, thus names should be separated into parts and the corresponding parts should be found. Splitting full names into parts without knowing the context (language, naming customs of the person, etc.) without building a dictionary of all names is a hard problem.

Using name similarities is insufficient to assess whether two physical persons having the same name represent the same person or not. Other attributes can be used to solve this problem and to increase the accuracy of the matching overall. Moreover, the names that have a high frequency in a population like *Jean Dupont* in French, *Alex Jones* in English,*Weiten Li* in Chinese, *Alexander Ivanov* in Russian or *Aarav Patel* in Hindi generate massive amount of false positive in the name matching process.

In this paper, we explore the current approaches of name matching and examine their

validity. Then we discuss limitations of such approaches and introduce several techniques which can be used to overcome the limitations in the multilingual name matching. Later we show how an existing model can be improved with the new proposed techniques to do the multilingual entity resolution and present the performance.

This paper enriches the literature related to name matching methods applied to KYC. We discuss limitations of such approaches and introduce several techniques which can be used to overcome the limitations in the multilingual name matching. Later we show how an existing model can be improved with Bayesian search theory(Eisenstein et al. (2011), Sadinle (2017)) to reduce the false positives in the multilingual entity resolution and present the performance. The paper is organized as follows: *Section 2* discusses reviews the different strategies used for Entity resolution, *Section 3* focuses on the methods for assessing names similarities in different language, *Section 4* describes the resolution model framework introducing the Bayesian approach for matching estimation , *Section 5* presents the results of name matching methods applied for data from the British and Russian business registries. *Section 6* concludes.

2. Entity resolution methods

Entity resolution methods can be categorized by the level of using relational information in matching, as described in Bhattacharya and Getoor (2007):

- Attribute-only Entity Resolution. Record similarity depends only on the similarities of the attributes.
- Naive Relational Entity Resolution. Record similarity depends on the similarities of the attributes of the two entities (as in the previous case) and the similarities of the attributes of the records related to the two records being matched.
- Collective Relational Entity Resolution. Record similarity depends on the similarities of the attributes of the two records (as in the attribute-only case) and the similarity of the records related to the two records being matched.

All of the categories from above require an approach to estimate the similarity between the attributes of two entities can use different. For datasets concerning companies and persons the main attribute of an entity is the name. Thus, name maching is the key aspect of the entity resolution applied to KYC process. Given a name represented by the string A in one language and a name represented by the string B in possibly other language, a name

matching algorithm should tell if *A* and *B* represent the same person or give a probability of this (Peng et al. (2018); Patman and Thompson (2003)).

One alternative to exact string matching is to convert name strings to some common phonetic representation of the names and then to compare the phonetic representations. Some of the possible phonetic representations are Soundex Russell (1918), Match Rating Approach Moore (1977), Daitch-Mokotoff Soundex Mokotoff (2007), Beider-Morse Phonetic Matching Beider (2008), Double Metaphone Philips (2000).

- Soundex algorithm Russell (1918) is the ancestor of phonetic name matching algorithms. Soundex maps names to a special code consisting of a letter and three digits. The letter is the first letter of the name and the digits describes approximately the consonants of the name. The initial aim of Soundex was to be easily computed manually by human and was designed to be applied to paper documents. Its performance is relatively poor compared to more recent development.

- Match Rating Approach (MRA) Moore (1977) employs a very basic but straightforward process that transforms the string by deleting the vowels if the names does not start with a vowel and by deleting the second consonants in the pairs of double consonants, thereby reducing the name to a maximum of six characters by retaining the first and the last three characters. After the two names are encoded the MRA gives the matching rating of the two names based on the similarity of the two strings.

- Daitch-Mokotoff Soundex Mokotoff (2007) is an alternation of the original Soundex for Yiddish and Slavic languages. In this algorithm names are given six digit codes. The first letter is coded too (contrast to original Soundex where the first letter was retained as it was). Names can have multiple codes which is different from the original Soundex where names are mapped to only one code.

- Beider-Morse Phonetic Matching (BMPM) Beider (2008) was designed to decrease the number of false hits produced by Soundex-like algorithms. BMPM incorporates more than 300 common rules and has a number of language-specific rules to support 10 languages. The first step of BMPM is to identify the language and only then the approximate phonetic value is calculated based on the language detected.

After the names were converted to their phonetic representations (or if no conversion was done and the names remained as they were given), the two strings should be compared with a string distance metric like Levenshtein distance, Guth algorithm, Jaro Cohen et al. (2003), Jaro-Winkler Winkler (2006), etc.

- 0IZIRWLXIMR HMWXERGI FIX[IIR X[S WXVMRKW MW E QIXVMG MRK XLI RYQFIV SJ SRI GLEVEGXIV GLERKIW WYFWXMXYXI E G VIQSZI E GLEVEGXIV MRWIVX E GLEVEGXIV RIIHIH XS GLERKI
- +YXL EPKSVMXLQ [EW WTIGM¼GEPP] HIWMKRIH XS GSQTEVI R ER MRTYX ERH EW ER SYXTYX TVSZMHIW [MXL XLI TVSFEFMPMX WTIPPMRK SJ SRI REQI +YXL…W EPKSVMXLQ GSQTEVIW X[S W WSQIXMQIW WOMTTMRK SV FEGOXVEGOMRK SRI SV X[S GLEVE

'SQTYXMRK 2EQI 7MQMPEVMXMIW FIX[IIR)RKPMWL ERH 6YWWM

-R XLMW WIGXMSR [I WTIGM¼GEPP] HMWGYWW QYPXMPMRKYE WMER ERH)RKPMWL REQIW -X WLSYPH FI RSXIH XLEX [I WTIGM¼G PERKYEKIW FIGEYWI SJ WIZIVEP VIEWSRW 8LI QEMR VIEWSR MW LMKLP] TLSRIXMG WTIPPMRK VITVIWIRXW TVSRYRGMEXMSR ERH RIXMG WTIPPMRK VITVIWIRXW TVSRYRGMEXMSR TSSVP] ERH XL IWTIGMEPP] [MXL 0EXMR EPTLEFIX EVI YWIH MR)RKPMWL MR XL VEXLIV XLER FIMRK XVERWJSVQIH XS FIXXIV VITVIWIRX XLI SVMK QSRX *VS]H %W E GSRWIUYIRGI SJ XLMW [I EVI EGXYEPP] LEZM REQI TEMVW FYX EPWS QER] SXLIV REQI TEMVW

8VERWPMXIVEXMSR

6YWWMER ERH)RKPMWL PERKYEKIW YWI HMJJIVIRX EPTLEFIXW [LMGL QEOIW WXVMRK WMQMPEVMX] JYRGXMSRW PMOI .EVS ;MR TEVMRK REQIW JVSQ XLIWI PERKYEKIW 8S SZIVGSQI XLMW MWWY 0EXMR WGVMTX

-R KIRIVEP JSV E REQI XLIVI GER FI WIZIVEP TSWWMFPI ZEPMH 7YGL GSRZIVWMSR GER FI ... MXIVEXM MW E QSVI W]WXIQEXMG ERH VIZIVWMFPI TVSGIHYVI ¹òþøðô XS)P RIXMG GSRZIVWMSR JSGYWIH SR TVIWIVZMRK XLI TVSRYRGMEXM XVERWPEXMSR MW HSRI ZME QETTMRK E REQI JVSQ XLI ¼VWX PE REQI SJ XLI WIGSRH PERKYEKI ¼ÃöÃòþ XS 2EXLEP] MJ XLIVI MW LEZI HMJJIVIRX VYPIW ERH GYWXSQW 0SV HMJJIVIRX PERKYEKI TE
(MJJIVIRX HSQEMRW GER GSRXEMR REQIW TVSHYGIH F] HMJJIV WMER MRXIVREXMSREP TEWWTSVXW LEZI REQIW GSRZIVXIH JVSQ WIX SJ XVERWPMXIVEXMSR VYPIW [LMGL EVI WSQIXMQIW HMJJIVI MR PIWW JSVQEP GSRXI\XW

*MKYVI '] VMPPMG 0EXMR ERH 0EXMR '] VMPPMG REQI ZEVME

)ZIR XLSYKL XVERWGVMTXMSR GER FI YWIH QSVI [MHIP] [I[MPP
IVEXMSR VYPIW JSV XLI MRXIVREXMSREP TEWWTSVXW FIGEYWI X

%PMKRMRK 2EQIW

2EQIW MR HMJJIVIRX PERKYEKIW ERH GYPXYVIW GER LEZI HMJJ
I K :PEHMQMV :PEHMQMVSZMGL 4YXMR LEZI E KMZIR REQI :PEH
ERH E JEQMP] REQI 4YXMR 7TERMWL REQIW I K 4IHVS 7›RGL
KMZIR REQI 4IHVS ERH X[S JEQMP] REQIW 7›RGLI^ ERH 4£VI^ 'E
I K 8LIVIWE 1EV] 1E] GSRWMWX SJ XLI ¼VWX REQI 8LIVIWE XL
XMSREP QMHHPI REQIW 1EV] 2EQI ZEVMEXMSR GER FI IRXEMPIH
REQI MR HMJJIVIRX PERKYEKIW
1SVISZIV E REQI GER ETTIEV MR HMJJIVIRX JSVQW REQI TEVX
:PEHMQMVSZMGL 4YXMR :PEHMQMV 4YXMR REQI TEVXW GER FI
QEXMGEP XVERWJSVQEXMSRW PMOI MR½IGXMSR :PEHMQMVE 4Y
8LMW ZEVMIX] SJ JSVQW QEOIW I\XVEGXMRK REQI TEVXW ERH G
LEVH XEWO 8S TEVXMEPP] SZIVGSQI X LPMW KRQMSRXIH[YMEXVESFYG
SJ GSQTYXMRK REQI WMQMPEVMX]
*YPP REQI WMQMPEVMX] SJ X[S REQIW [MXL EPMKRQIRX MW HI½
SJ EPP TEVX TIVQYXEXMSRW : 4YXMR 4YXMR : [MXL XLI WIGSR
X[IIR :PEHMQMV 4YXMR ERH 4YXMRE 0]YHQMPE [MPP FI LMKL FYX

GSQTYXI TIVQYXEXMSRW SJ XLI WIGSRH REQI EWWYQMRK XLI WM

4LSRIXMG 8VERWJSVQEXMSRW

'SRZIVXMRK REQIW FIX[IIR PERKYEKIW GER QEOI REQIW PIWW V [LIR XVERWPEXMSRW SV LMWXSVMGEP XVERWPMXIVEXMSR GYWX XIV ˆ,‰ WS YRH XVERWGVMFIH JVSQ +IVQER XS 6YWWMER ERH XLIR XS)RKPMWL QMKLX GSRZIVXMRK ,IVQERR XS 6YWWMER ERH XLIR XS)RKPMWL QMKLX SRP] I\EQTPI SJ WYGL MRGSRWMWXIRGMIW %RSXLIV I\EQTPI MW MW ORS[R EW 6³WXIQ MR 8YVOMWL 6¯WXžQ MR 8EXEV ERH ¿ÉÈöÄ MRK XS)RKPMWL F] HVSTTMRK HMEGVMXMGW ...

6YWXIQ 6YWXEQ

7SQIXMQIW REQIW LEZI IUYMZEPIRXW JSV REQIW MR SXLIV PERK REQI »ðÊÃðò 1MOLEMP LEW E 9OVEMRMER GSYRXIVTEVX »ðÊÃÅ REQIW [SYPH FI TVMRXIH MR XLI TEWWTSVX MR 7SZMIX 9RMSR ERH FI 1MOLEMP FYX RS[EHE]W XLI REQI MW WTIPPIH EW 1]OLEMPS

-R XLI PEWX WIGXMSR SJ XLI TETIV [I YWI 4LMPMTW (SYFPI 1IXETLSRI REQIW (SYFPI 1IXETLSRI [EW HIWMKRIH XS EGGSYRX JSV XLI HMJJ MR HMJJIVIRX PERKYEKIW *SV I\EQTPI (SYFPI 1IXETLSRI GSHIW JVSQ »ÃÅñò [MPP FI IUYEP XS 1/0

)RXMX] 6IWSPYXMSR 1SHIP (IWGVMTXMSR

*SPPS[MRK XLI JSVQEPMWQ MR XS HIJMRI E]/SYRK VSHIP GI XLI 4VSFEFMPMWXMG 7SJX 0SKMG J4VEQI[SVO 470 XS HIWG 2 [0;1] ERH IPE\EXMSR SJ XLI PSKMGEP QSHIPW 9RHIV XLI VIPE\E QMK IX EP 470 HIVMZIW XLI SFNIGXMZI JYRGXMSR F] XVERWPEX HITIRHIRGMIW FIX[IIR ZEVMEFPIW ERH IZMHIRGI MRXS LMRKI PSW XVERWPEXMSR F] YWMRK XLI 0YOEWMI[MG^ RSVQ ERH GS RSVQ X PSKMGEP GSRRIGXMZIW

$$p \wedge q = \max(0; p + q - 1)$$

$$p \vee q = \min(1; p + q)$$

$$\neg p = 1 - p$$

%W E VIWYPX SJ XVEMRMRK XLI QSHIP [MPP FI EFPI XS HIXIVQMRI XLI KMZI b VIJIVVMRK XS XLI WEQI TIVWSR SV XLI WEQI IRXMX] SameAs(a; b)

8LI TVSTSWIH QSHIP GSRWMWXW SJ E RYQFIV SJ 470 VYPIW 8LI
MRXS XS[GEXIKSVMIW EXXVMFYXI WMQMPEVMX] VYPIW ERH VIPE
%XXVMFYXI WMQMPEVMX] VYPIW WXEXI XLEX MJ X[S QI EXXVMFYXI MW WMQMPEV M
VIJIVIRGIW WLSYPH FI QEXGLIH ERH MJ WSQI EXXVMFYXI MW RSX
XLI VIJIVIRGIW WLSYPH RSX FI QEXGLIH *SV I\EQTPI JSV X[S TIVW
a ERH b [I HI¼RI

$$SimName_{JW}(a;b) \Rightarrow Same(a;b)$$
$$\neg SimName_{JW}(a;b) \Rightarrow \neg Same(a;b)$$
$$SimDOB(a;b) \Rightarrow Same(a;b)$$
$$\neg SimDOB(a;b) \Rightarrow \neg Same(a;b)$$
$$SameGender(a;b) \not\Rightarrow Same(a;b)$$
$$\neg SameGender(a;b) \Rightarrow \neg Same(a;b)$$

;LIVI $SimName_{JW}$ MW XLI .EVS ;MROPIV REQI WMQMPEVMX] $SimDOB$ MW XLI X[S V
¼RIH EW QE\MQYQ SJ XLI HEXIW SJ FMVXLW HMZMHIH F] XLI QMRM
$SameGender(a;b)$ MW E FMREV] SFWIVZIH EXSQ XLEX XEOIW MXW ZEPYI JV
$a.gender = b.gender$ $Same(a;b)$ MW E GSRXMRYSYW ZEPYI XS FI MRJIVVIH [LM
EFMPMX] XLEX a ERH b VIJIV XS XLI WEQI TIVWSR 8S MPPYWXVEXI 470 MR E
GSRXI\X XLI JSPPS[MRK VYPI IRGSHIW XLEX QIRXMSRW [MXL WM
WEQI KIRHIV QMKLX FI XLI WEQI TIVWSR

$$SimName(a;b) \wedge SimDOB(a;b)\ SameGender(a;b) \Rightarrow Same(a;b)$$

6IPEXMSREP EXXVMFYXI VYPIW GSQFMRI XLIV [MXL XLI EXXVMFYXI WMQM
EXXVMFYXI VYPIW WXEXI XLEX MJ X[S VIJIVIRGIW QIRXMSR WMQM
XSVW [MXL WMQMPEV REQIW FSXL LEZI XLI WEQI EHHVIWW FSXL
XLIR XLI X[S VIJIVIRGIW WLSYPH FI QEXGLIH x ERH y
EXXVMFYXIW VYPIW GER FI

$$HasDirector(x) \wedge HasDirector(y) \wedge SimDirector(x;y) \Rightarrow Same(x;y)$$
$$HasAddress(y) \wedge HasAddress(y) \wedge SimAddress(x;y) \Rightarrow Same(x;y)$$

[LIVI $SimDirector$ ERH $SimAddress$ EVI HI¼RIH EW XLI QE\MQYQ SJ 0IZIRWLXIM
;MROPIV WMQMPEVMXMIW SJ HMVIGXSVW… REQIW ERH EHHVIWWI
0IX…W EWWYQI JSV I\EQTPI XLEX [I [SYPH PMOI XS EWWIWW XLI F
TERMIW % JVSQ E FYWMRIWW VIKMWXV] MI 9RMXIH /MRKHSQ E

JVSQ ERSXLIV FYWMRIWW VIKMWXV] MI 6YMWW E [M*VISSWSE WIPSE XM
SimDirector ERH X[S QIRXMS R WH X[S A = fx : t(; x)g ERH B = fy : t(; y)g
[LMGL EVI XLI WIXW SJ QIRXMSRW W TFSXM ZXIS] [MXL XLI WMPEXMSR
SimDirector A ERH [MPP FI VITVIWIRXIH F] XLI WIX EW SVIEPT (MXIGIRSVW
8LYW XLI WMQMPEVMX] ERH [MPP FI XLI VIPEXMSREP EXXVMFYX
ERHERH MW HI¼RIH EW

$$\mathrm{Sim}(A;B) = \frac{1}{|A|}\sum_{x \in A} \max_{y \in B} \mathrm{SimName}(x;y)$$

;I GER WII XLEX XLI REQI WMQMPEVMXMIW IWXMQEXIW EVI YWII
QSHIP XLYW XLI TIVJSVQERGI SJ QEXGLMRK HITIRHW LIEZMP] SR
,S[IZIV XLI SVMKMREP SimName(x;y) SR XSM LIEW WSQI PMQMXEXMS
TP]MRK MX EW MW XS SXLIV HEXEWIXW
*MVWX XLI YRHIVP]MRK QIXVMGW GSQTYXI WMQMPEVMXMIW F
PEWX REQIW &SXL SJ XLIWI JYRGXMSRW EVI PIJX YRHI¼RIH JSV X
VIWIRXIH EW SRI WXVMRK ERH MX MW YRORS[R LS[XS WITEVEXI R
REQI IXG
7IGSRH .EVS ;MROPIV ERH 0IZIRWLXIMR HMWXERGIW [LMGL [M
WXVMRKW [VMXXIR MR X[S HMJJIVIRX EPTLEFIXW I K 0EXMR ERH
GEFMPMX] SJ XLI ETTVSEGL MR QYPXM PMRKYEP GSRXI\X ERH MR
IVEXMSR QSHIP YWIH
8LMVH JSV REQIW LEZMRK ER MRGVIEWI JVIUYIRG] MR E TSTY
IQTPS]IH MR XLI JVEQI[SVO EFSZI [MPP KIRIVEXI E LMKL RYQFIV SJ
SVHIV XS VIHYGI XLMW FMEW MX MW RIGIWWEV] XS XEOI MRXS EGGS
SJ E REQI MR XLI [E] XLI WMQMPEVMX] JYRGXMSRW MW FYMPX
;LIR WIEVGLMRK JSV XLI GSVVIWTSRHMRK IRXMXMIW [I RSVQEPM
FIEPXIV MR SVHIV XS XEOI MRXS EGGSYRX XLI ETVMSVM JVIUYIRG
XLI XSXEP TSTYPEXMSR [LMGL WIVZIW EW F_b [MXL XLI WMQMPEVMX]

$$\mathrm{SimName}_{JW}(a;b) = \sum_{b_i \in P} \mathrm{SimName}_{JW}(a;b) \cdot f(b;F_b)$$

HITIRHMRK SR E &E]IWMER $f(b;F_b)$ [LMGL MW XLI RSVQEPM^MRK XLI REQI
MRGVIEWI JVIUYIRG] MR XLI TSTYPEXMSR

%TTPMGEXMSR XS FYWMRIWW VIKMWXVMIW HEXE JVSQ 6YWWM

(EXEWIX TVIWIRXEXMSR

-R SVHIV XS EWWIWW XLI I¾GMIRG] SJ XLI ZEVMSYW IRXMX] VIW
VEG] SJ XLI REQIQEXGLMRK QIXLSHW [I FYMPX E VIPEXMSREP HEX
?iiTb,ff#2i X+QKT MB2b?Qm EIR2H6YWWmTE,Rf2;`mHXM HQ;X`mfBM/2
?iKHFYWMRIWW VIKMWXVMIW 8LI HEXEFEWIW GSRXEMR MRJSVQ
XMSR EFSYX XLI GSQTERMIW... OI] TIVWSRW ERH VIPEXMSRWLMTV
XLI GSQTER] MI HMVIGXSV QEREKIV

8LI HIWGVMTXMZI TVIWIRXEXMSR SJ XLI HEX8LWMXRMSWVIQEXMSR
EFSYX 6YWWMER GSQTERMIW ERH XLIMV OI] TIVWSRW MW TVIWIR
XIR ']VMPPMG EPTLEFIX [LMPI MRJSVQEXMSR EFSYX &VMXMWL G
)RKPMWL [VMXXIR MR 0EXMR EPTLEFIX

6IKMWXIV	2S SJ YRMZ2SYISJ YRMZ2SYISJ VIPEX%WXRVMFYXIW HMVIGXSVW TIVWSRW GSQTERMIWWLMTW	%XX
9RMXIH HSQ	/MRK XMSREPMX]	2EQI (3& %HH2EQIW %HEHVIWW 8 FIV HEXI SJ MRGS
6YWWMER IVEXMSR	*IH	2EQI 8E\RYQFIV 2EQI %HHV FIV HEXI SJ MRGS

8EFPI 4VIWIRXEXMSR SJ HEXEWIXW YWIH MR XIWXMRK IRXMX] VIWS

+MZIR XLI WIX SJ OI] TIVWSR MR 6YWWMER GSQTERMIW [I XV] XS
TIVWSRW MR &VMXMWL GSQTERMIW 8LMW X]TI SJ EREP]WMW [SY
HIEPMRK JSV I\EQTPI [MXL GPMIRXW XLEX GSYPH FI MRZSPZIH MR
MRK WERGXMSR 8LI WIEVGL MW IQTPS]MRK XLI QIXLSHW HIWGVM
WXITW

- (EXE GPIERMRK MRGPYHMRK IPMQMREXMSR SJ MRGSQTPIXI F
- &EWMG XI\X TVSGIWWMRK I\GPYWMSR SJ TYRGXYEXMSR SV X
- 8VERWPMXIVEXMSR SJ XLI TIVWSRW ERH GSQTERMIW... REQIW
- 2EQI QEXGLMRK YWMRK ZEVMSYW WMQMPEVMX] JYRGXMSR QI
-)RXMX] VIWSPYXMSR FEWIH SR ZEVMSYW WXVEXIKMIW

)RXMX] VIWSPYXMSRW QIXLSHW ETTPMIH

1EXGLMRK TIVJSVQERGIW EVI QIEWYVIH [MXL E WIX SJ HMJJIVI
PYXMSR ERH E WIX SJ EPKSVMXLQW XS 1YWXSV¼ REXIEP *SV XLIK IRXMX]
VIWSPYXMSR XLI WXVEXIKMIW YWMRK 470 PSKMG EVI GSRWMHIVI

2EQIGSRWMXMRK MR WMQTPI REQI QEXGLMRK

2EQIW 4IVWSREP -RJS 4- 6IPEXMSREP -RJS 4- QEXGLMRK MR REQI QEXGLMRK ERH QEX
MRJSVQEXMSR %HHVIWW +IRHIV (EXI SJ &MVXL

2EQIW 4IVWSREP -RJS 4- 6IPEXMSREP -RJS 6- WMQMPEVMX] WXVMRK
REQI QEXGLMRK QEXGLMRK SJ TIVWSREP MRJSVQEXMSR ERH
REQIW

8LI IUYMZEPIRGI SJ JYPP REQIW I K ¼VWX REQI ERH PEWX R
XLI OI] WXIT SJ EPP IRXMX] VIWSPYXMSR WXVEXIKMIW EVI EWWIW
:EVMSYW QIXLSHW EVI IQTPS]IH

-R 8VERWPMXIVEXMSR QEXGLMRK REQI WMQMPEVMXMIW [IVI GSQTYXIH EW WM
ERH 6YWWMER ZIVWMSRW SJ XLI REQIW 8VERWPMXIVEXMSR [EW TIVJSVQIH F
F] XLI EPMKRQIRX TVSGIHYVI HIWGVMFIH MR 8VERWPMXIVEXMSR %PMKR
8VERWPMXIVEXMSR MW XLIR IRLERGIH F] TLSRIXMG QEXGLMRK
XLI VIWYPXW EVI WLS[R MR 8VERWPMXIVEXMSR 4LSRIXMG

8[S WMQMPEVMX] JYRGXMSR [IVI YWIH XLI GPEWWMG .EVS ;MR
QIXVMG [MXL E &E]IWMER GSVVIGXMSR

8EFPI ERH WLS[XLI JVIUYIRG] SJ PEWX ERH ¼VWX REQIW MR X
TIVWSRW JSV &VMXMWL ERH 6YWWMER GSQTERMIW VIWTIGXMZI
XLIVI MW MRGVIEWI PMOIPMLSSH SJ QEXGLMRK XS E WMRKPI] REQIW
[VSRKP] VIWSPZI XLI IRXMX] HYI XS LMKL SGGYVVIRGI SJ XLMW ¼V
-R XLI GEWI SJ SYV WIEVGL PSSOMRK JSV OI] TIVWSR JVSQ 6YWWM
MWXV] MX GSYPH FI XLEX XLI REQIW MR 6YWWMER EVI XVERWPMXIVEXI JEPWI TSWMX
*SV XLMW VIEWSR XLI &E]IWMER GSVVIGXMSR MW RIGIWWEV] MR S

6IWYPXW

8LI JSPPS[MRK QIXVMGW EVI YWIH MR SVHIV XS EWWIWW XLI S
VMXLQW

- 4VIGMWMSR XLI VEXMS SJ XLI XVYI QEXGLIW JSYRH XS EPP JSYRH
  XLI JSYRH QEXGLIW EVI XVYI QEXGLIW

0EWX REQIW			*MVWX REQI		
2EQI	2YQFIV	*VIUYIRGE	2EQI	2YQFIV	*VIUYIRG]
71-8,	35317	080	(%:-(	130799	296
.32)7	27319	062	.3,2	111393	252
&63;2	19580	044	1-',%)0	88622	200
;-00-%17	19560	044	4%90	78235	177
8%=036	17769	040	%2(6);	76391	173
(%:-)7	15631	035	4)8)6	70445	159
4%8)0	14225	032	63&)68	63055	142
;-0732	12934	029	6-',%6(	62991	142
8,31%7	12459	028	.%1)7	62316	141
):%27	12234	028	1%6/	60208	136
/,%2	11920	027	78)4,)2	58922	133
.3,2732	11025	025	',6-7834,)6	57216	129
7-2+,	10080	023	-%2	41098	093
63&)687	9487	021	7-132	38681	087
63&-2732	9373	021	;-00-%1	34118	077
;%0/)6	9329	021	%28,32=	34036	077
8,31473 2	9199	021	2-',30%7	30127	068
;,-8)	9175	021	(%2-)0	29103	066
,%00	9062	020	.32%8,%2	29092	066
,%66-7	9035	020	1%68-2	29076	066

8EFPI *VIUYIRG] SJ PEWX ERH ¼VWX REQIW MR XLI WEQTPI SJ XLI OI] TIVW

- 6IGEMPW7 XLI VEXMS SJ XLI XVYI QEXGLIW JSYRH XS EPP XVYI QE X[WIXW M I [LEX TEVX SJ EPP QEXGLIW [EW JSYRH
- * WGSVIW XLI LEVQSRMG QIER SJ TVIGMWMSR ERH VIGEPP

MW XLI [SVWX TSWWMFPI TVIGMWMSR VIGEPP * WGSVI ERH MV 8LI TIVJSVQERGI SJ XLI QSHIP SR HMJJIVIRX HEXE [MXL HMJJIVIP MW WLS[R MR

(MWGYWWMSR

8LI VIWYPXW WLS[XLEX XLI ETTVSEGL YWMRK XVERWPMXIVEXM XIGLRMUYIW TVSZMHIW KPSFEPP] [MXL FIWX VIWYPXW 8LI &E]IW XVMFYXMSR SJ REQI JVIUYIRG] MQTVSZIW QEWWMZIP] XLI TVIGMV RIKEXMZIW 7IZIVEP MRWMKLXW GER FI HIVMZIH JVSQ XLIWI VIWY %PMKRMRK ERH 4LSRIXMG XVERWJSVQEXMSR MQT%8ZMIWR WPMI MRK TVSGIHYVI HIWEGRVHMFMHQTVSZIW VIGEPP SJ QEXGLMRK F] VISV

0EWX REQIW			*MVWX REQI		
2EQI	2YQFIV	*VIUYIRG]	2EQI	2YQFIV	*VIUYIRG]
-:%23:			%0)/7%2(6		
/9>2)873:			7)6+)=		
4343:			:0%(-1-6		
71-623:			%2(6)=		
-:%23:%			%0)/7)=		
4)863:			(1-86-=		
:%7-0=):			=)02%		
/9>2)873:%			2-/30%=		
23:-/3:			8%8=%2%		
1-/,%=03:			=):+)2-=		
4%:03:			=96-=		
73/303:			-+36		
1363>3:			2%8%0=%		
/3>03:			1-/,%-0		
:30/3:			30+%		
78)4%23:			:-/836		
1%/%63:			30)+		
*)(363:			-6-2%		
7)1)23:			7:)80%2%		
=)+363:			:%0)6-=		

8EFPI *VIUYIRG] SJ PEWX ERH ¼VWX REQIW MR XLI WEQTPI SJ XLI OI] TIVWS

REQIW XS MRGVIEWI XLI WMQMPEVMX] -J XLI ¼VWX REQI MW QIR
3PIK XLMW MW EPWS XEOIR MRXS EGGSYRX GSQTEVIH XS XLI FEW
MRK GER MRGVIEWI XLI WMQMPEVMX] SJ EGXYEPP] RSR QEXGLMR
SJ QEXGLMRK

8LMW MQTVSZIQIRX IRGSQTEWWIW ZEVMEXMSRW SJ XLI XVERW
XMSRW EFSZI *S3VPIKE1QSTVPSWVZGSVIGPX] QEXGLIH EKEMRWX XLI 0
3PIKW 1SVS^SZW

9WMRK (SYFPI 1IXETLSRI GSHIW JSV GSQTYXMRK WMQMPEVMX
TVIGMWMSR SJ QEXGLMRK F] TYXXMRK WIZIVEP HMJJIVIRX XVER

*IEXYVI 7IX	4VIGM WMSR	6IGEPP	7GSVI
8VERWPMX 2EQIW			
8VERWPMX 2EQIW %PMKR			
8VERWPMX 2EQIW %PMKR 4LSRIXMG			
8VERWPMX 2EQIW 4-			
8VERWPMX 2EQIW %PMKR 4-			
8VERWPMX 2EQIW %PMKR 4LSRIXMG 4-			
8VERWPMX 2EQIW 4- 6			
8VERWPMX 2EQIW %PMKR 4- 6			
8VERWPMX 2EQIW %PMKR 4LSRIXMG 4- 6			

8EFPI 4IVJSVQERGI JSV ZEVMSYW WXVEXIKMIW SJ IRXMX] VIWSPYXMSR ERH Z
SR GPEWWMG WMQMPEVMXMIW QIXVMG

*IEXYVI 7IX	4VIGM WMSR	6IGEPP	7GSVI
8VERWPMX 2EQIW			
8VERWPMX 2EQIW %PMKR			
8VERWPMX 2EQIW %PMKR 4LSRIXMG			
8VERWPMX 2EQIW 4-			
8VERWPMX 2EQIW %PMKR 4-			
8VERWPMX 2EQIW %PMKR 4LSRIXMG 4-			
8VERWPMX 2EQIW 4- 6			
8VERWPMX 2EQIW %PMKR 4- 6			
8VERWPMX 2EQIW %PMKR 4LSRIXMG 4- 6			

8EFPI 4IVJSVQERGI JSV ZEVMSYW WXVEXIKMIW SJ IRXMX] VIWSPYXMSR ERH Z
SR WMQMPEVMXMIW QIXVMG [MXL &E]IWMER GSVVIGXMSR

FYGOIXI K *SV%PPEQTPIRHIVERHE3XVIIRWHPM XIVEXMSR SJ ¶òÄñÈÃô
XLI WEQI (SYFPI 1IXETLSRI [LMGL VIWYPXW MR XLI QE\MQYQ REQI
%GGSYRXMRK JSV TIVWSREP MRJSVQEXMSR EW ER IRXMX] VIWSI

TVIGMWMSR E4RMWVSGREPPMRJSVQEXMSR MW EZEMPEFPI SRP] JSV
WSRW ERH PIWW EZEMPEFPI JSV 6YWWMER HEXE &YX MR QER] GE
MQTPMIH JVSQ XLI XMXPI 1V ZW 1VW MR 9/ ERH JSV XLI XIVQMRE
IRXMXMIW SZ ZW SZE

%GGSYRXMRK JSV VIPEXMSRWLMT MRJSVQEXMSR EW ER IRXMX
WMZIP] XLI TVIGMWMSR ERH F%&GVSHYRXMWRXLJSVVGEPEXMSRWLMT
PSSO JSV TISTPI [MXL WMQMPEV REQIW XLEX EVI MRZSPZIH MR
&VMXMWL 4IXVSP 08(ERH 333 &VMXMWL 4IXVSP 8LMW VIHYGIW
XMZIW FYX VIHYGIW XLI VIGEPP EW XLI QEXGLIW VITVIWIRX SRP]

*EPWI TSWMXMZIW EVI KIRIVEXIJSV RSR46YWSVWE[MXLMRKW
REQIW PMOI =IR 'LER 7IR SV 7ER &S 0M XLEX LEZI GSQTERMIW MR 6
[LIR QEXGLMRK TIVWSRW JVSQ XLI &VMXMWL VIKMWXV]

8LI &E]IWMER GSVVIGXMSR SJ XLI WMQMPE;LMXPIQPIXWWJVMQYTVS
REQIW EW 3PIK (IVMTEWOE EVI GSVVIGXP] QEXGLIH F] EPP EPKSVM
VSSX GEYWI SJ JEPWI TSWMXMZI 8LMW MW EPWS HYI XS XLI JEGX X
HS RSX VIKMWXIV W]WXIQEXMGEPP] XLI TEXVSR]QMG REQI [LMGL

'SRGPYWMSRW

8LMW TETIV I\TPSVIW QIXLSHW SJ QYPXMPMRKYEP IRXMX] QEX
REQI QEXGLMRK ETTVSEGL YWMRK WXVMRK GSQTEVMWSR QIXVM
ERH [MXL VIPEXMSREP MRJSVQEXMSR 8LI VIWYPXW WLS[LS[HM
IRXMX] VIWSPYXMSR WXVEXIKMIW EJJIGX TVIGMWMSR ERH VIGEP
GPEWWMGEP WMQMPEVMX] QIXVMGW MR SVHIV XS EGGSYRX JSV X
TVSZMHIW [MXL FIXXIV TVIGMWMSR 2IZIVXLIPIWW XLI TVIWIRXIP
XLI MWWYIW SJ REQI ZEVMEFMPMX] MR XLI XVERWPMXIVEXMSR XV
PERKYEKIW [MXL HMJJIVIRX EPTLEFIXW 3YV VIWYPXW LMKLPMKL
F] VITPEGMRK XLI WXEXMG ERH SRI HMQIRWMSR GSRGITX SJ /RS[=
QYPXM HMQIRWMSREP ERH JSV[EVH PSSOMRK GSRGITX SJ ‰/RS[
[MPP FI I\TPSVIH MR E JYVXLIV VIWIEVGL

6IJIVIRGIW

&EGL 7 , &VSIGLIPIV 1 ,YERK & +IXSSV 0 ,MRKI PSWW
TVSFEFMPMWXMG WSJX PSKMG EV<MZ TVITVMRX EV<MZ

&IMHIV % &IMHIV QSVWI TLSRIXMG QEXGLMRK %R EPXIVRE
LMXW %ZSXE]RY XLI-RXIVREXMSREP 6IZMI[SJ .I[MWL +IRIEPS

&LEXXEGLEV]E - +IXSSV 0 'SPPIGXMZI IRXMX] VIWSPYXM
EGXMSRW SR /RS[PIHKI(MWGSZIV] JVSQ(EXE 8/((

'LVMWXIR 4 (EXE QEXGLMRK GSRGITXW ERH XIGLRMUYIW
PYXMSR ERH HYTPMGEXI HIXIGXMSR 7TVMRKIV 7GMIRGI &YWI

'SLIR ; 6EZMOYQEV 4 *MIRFIVK 7 % GSQTEVMWSR SJ V
MRK REQIW ERH VIGSVHW MR /HH[SVOWLST SR HEXE GPIERMR
•

)MWIRWXIMR . =ERS 8 'SLIR ;; 7QMXL 2 % <MRK) 4
REQIH IRXMXMIW JVSQ FE]IWMER RSRTEVEQIXVMGW MR 4VSGI
9RWYTIVZMWIH 0IEVRMRK MR 204 %WWSGMEXMSR JSV 'SQTYXE

/MQQMK % &EGL 7 &VSIGLIPIV 1 ,YERK & +IXSSV 0 %
EFMPMWXMG WSJX PSKMG MR 4VSGIIHMRKW SJ XLI 2-47 ;SVOW
*SYRHEXMSRW ERH %TTPMGEXMSRW TT •

/SYOM 4 4YNEVE . 1EVGYQ ' /SILP] 0 +IXSSV 0 'SPP
JEQMPMEP RIX[SVOW MR (EXE 1MRMRK -'(1 -))) -RXIVREX
TT •

0IZIRWLXIMR : &MREV] GSHIW GETEFPI SJ GSVVIGXMRK WT
SJ SRIW 4VSFPIQW SJ MRJSVQEXMSR 8VERWQMWWMSR •

0M ' ; ' *SVIMKR REQIW MRXS REXMZI XSRKYIW ,S[XS XVE
KYEKIW,XVERWPMXIVEXMSR TLSRSPSKMGEP XVERWPEXMSR R
PEXMSR XLISV] 8EVKIX -RXIVREXMSREP .SYVREP SJ 8VERWPEX

1SOSXSJJ + 7SYRHI\MRK ERH KIRIEPSK]
LXXT [[[EZSXE]RY GSQ WSYRHI\ LXQP

1SSVI + & %GGIWWMRK MRHMZMHYEP VIGSVHW JVSQ TIVW
MHIRXM¼IVW ZSPYQI 97(ITEVXQIRX SJ 'SQQIVGI 2EXMSREP

1YWXE¼R - *VYR^E 1 ' 0II . 1YPXMPMRKYEP IRXMX]QE
JIVIRGISR%HZERGIH-RJSVQEXMSR2IX[SVOMRK ERH%TTPMGE

4EXQER * 8LSQTWSR 4 2EQIW % RI[JVSRXMIV MR XI\X
'SRJIVIRGISR-RXIPPMKIRGI ERH 7IGYVMX]-RJSVQEXMGW 7TVM

4IRK 8 0M 0 /IRRIH] . %GSQTEVMWSR SJ XIGLRMUYIW J
.SYVREP SR'SQTYXMRK .S'

4LMPMTW 0 8LI HSYFPI QIXETLSRI WIEVGL EPKSVMXLQ ' '

6YWWIPP 6 -RHI\b,9f0rX;QQ;H2X+QKfT i2MibYf7aHEXeIRXed

7EHMRPI 1 &E]IWMER IWXMQEXMSR SJ FMTEVXMXI QEXGLM
XLI%QIVMGER 7XEXMWXMGEP%WWSGMEXMSR •

7E•W * 4IVRIPPI 2 6SYWWIX 1 ' 0 V %PSKMGEP QIXLSH
MR 4VSG %%%- TT •

7MRKPE 4 (SQMRKSW 4)RXMX]VIWSPYXMSR[MXL QEVOS
-'(1… 7M\XL-RXIVREXMSREP'SRJIVIRGISR -))) TT •

;MROPIV ;) 8LI WXEXI SJ VIGSVH PMROEKI ERH GYVVIRX VIW
6IWIEVGL(MZMWMSR 97'IRWYW&YVIEY 'MXIWIIV

;MROPIV ;) 3ZIVZMI[SJ VIGSVH PMROEKI ERH GYVVIRX VIW
SJ XLI'IRWYW 'MXIWIIV

=IVQSPSZMGL (-QIRE WSFWXZIRR]]I RE WX]OI]E^]OSZ M OY
PERKYEKIW ERH GYPXYVIWA 1SWGS[6 :EPIRX

SCHWARZTHAL TECH

Know You Network, AI meets KYC

Visit schwarzthal.tech for more

Marius Frunza
marius.frunza@schwarzthal.com

Ilgiz Mustafin
ilgiz.mustafin@schwarzthal.com

Georgi Ormelyanov
georgi.ormelyanov@schwarzthal.com