Logo, company name

Description automatically generated

IntuView Name Matcher  
-  
User Guide

# Background

Identification, disambiguation and matching of names that are received in intelligence and law enforcement channels and linking of those names to entities in a watch list is a critical element of law enforcement, intelligence and counter-terrorism. Name matching is essential in vetting candidates for visas and political asylum, airline passenger manifests, matching of names received from human and intercept intelligence, and triage of open source material that may relate to people in a watch list.

The underlying problem of Name Analysis is the ambiguity of names that come from different societies and ethnic groups and are written indigenously in different orthographic systems. Names reflect cultural naming conventions, are received from different sources and are corrupted by different linguistic backgrounds of the transmitters and the transcribers of the names. The key obstacles for name vetting and matching is rooted in the fact that string matching yields low recall and phonetic matching yields low precision.

The representation of a name (transcription and identification of the name components such as given name, patronymic or family name is the outcome of various factors: the ethnic origin of the name, the occurrence of sounds in the name that the western ear is not used to and therefore mis-hears, the absence of letters in the Latin alphabet to represent certain letters of other languages, the literacy of the asylum seeker or person carrying the name, dialectical variants and more. Using IntuView Name Matcher guarantees that all names entered truly represent the original name and information are normalized across all the databases.

Names can yield information on an individual that is not explicit. Such information includes:

1. The individual’s religion or sect within the religion (names may be typically Muslim Sunni, Shiite, Jewish or Christian).
2. Honorifics that define the person’s status.
3. The role of an individual in his family/clan/tribe according to his lineage.
4. Socio-analysis of links and alliances between families through patronymics (fathers’ names as a name component of the offspring), matronymics (name components that relate to the person’s mother’s lineage) etc.
5. The generation or age of the individual – certain names may be more typical to certain age groups (for example: Ethel and Audrey are old names in the USA). In some cultures (Korea for example) an individual bears a generational name. This information can be collated from national statistics, school records etc.
6. Linguistic/cultural background of an individual or group - the name of a person reflects his or her cultural background and occasionally even the ideological leanings of his parents (e.g. a Syrian whose given name is Lenin on one hand or mujahid on the other hand).
7. Physical, occupational or ethnic background of an individual or his image in the eyes of his peers (through interpretation of nicknames). In Arabic, a person may be referred to as al-Masri (the Egyptian) if he is from Egypt, al-Tawil (the tall) if he is exceptionally tall etc. These are frequently misperceived as names.

# IntuScan Name Matcher Application

IntuView Name Analyser is a product of IntuView Ltd. Its purpose is to provide a non-intrusive, honest, precise and unbiased platform for correctly vetting, analyzing, normalizing, storing (and thus finding) and sharing names within and between agencies.

The Name Matcher allows the user to rapidly vet, validate, disambiguate, aggregate and match person names using cultural-based rules. The Name Matcher can be applied to any task that calls for vetting of names in order to alert the user if there may be cause for continued investigation. Some examples could include:

* Accurate and thorough matching of names against watch lists to reduce false positives and false negatives.
* Identifying anomalies where people’s apparent nationality does not match their passport (note that in many countries naturalization is not as common as in the United States).
* Flagging of potential familial or other affiliations between persons whose names are being checked and persons in a watch list.
* Flagging names of persons as potentially affiliated with high-risk places, tribes, clans or other non-obvious affiliations.
* Vetting watch lists to discover redundancies and duplicate entries of individuals, that once merged may yield important information.

IntuView Name Matcher currently includes extensive algorithms for names from the following ethnicity/language groups:

* Arabic names are fully supported including breakdown of ethnicity of the names to countries, regions (Maghreb, Levant, Gulf) and in some cases even tribal groups.
* Farsi, Afghani, Pakistani and Turkic names are supported at the level of identification of the ethnicity and resolution of the names by naming conventions.
* Indian names based on the most common Hindi naming conventions (Punjabi, Gujarati, Sikh, etc.). Indian Muslim names are identical in form to Pakistani names.
* Central Asian Muslim names (e.g. from Chechnya, Dagestan, Uzbekistan, Azerbaijan etc.) are identified as belonging to their general ethnic origin (Farsi, Turkic, Urdu). Names that are specific to these countries are identifies as such.
* Western names (Anglo-Saxon, French, Germanic, Hispanic, etc.) and Slavic names (Russian, Central Asian, Eastern European, Balkan) are identified as belonging to these groups and resolved through matching of naming convention, nicknames and variants of names. The next version of Name Matcher will apply back transliteration to Cyrillic based names.
* Asian names are identified and resolved in their canonic transliteration standards (Pinyin, Wade-Giles for Chinese; Hangul for Korean and Hepburn for Japanese). Back transliteration for these names is also planned.
* African names based on the most common tribal and ethnic naming conventions..

Upon input of new names, the Name Matcher provides the following:

* Identification of ethnicity of the input name (e.g. Anglo-Saxon, Hispanic, Arabic, Afghani, Iranian, etc.). In many cases, the identification will be more specific (e.g. Maghreb or North African or specifically Moroccan or Tunisian within that).
* Each name is analyzed to identify its constituent parts (given name, patronymic, family name, tribal name, nickname etc.). In many cases, the name may be written with the family name first. In language systems that the user is familiar with, this may be trivial (if we see Smith John, we understand that the more likely name is John Smith), however in foreign languages this is not transparent (Abdallah Muhammad can easily be Muhammad Abdallah since both are both given and family names). In such a case, the system will identify the correct order of the components.
* Name components are analyzed in order to extract implicit information such as national origin, gender, religion or sect identity etc. The name components then are linked to potential variants (e.g. a Romanized given name Ehsan may be a Moroccan variant of Hassan or the name Ihsan which is a different name). Each variant is kept as a potential identification of that name part. This is later used in order to facilitate matching with the variants of other names. All culturally acceptable variants of the parsed name are generated and validated with algorithms based on the naming conventions of the origin-culture of the name.
* Names and information input with them are examined to identify possible anomalies, inconsistencies, incompleteness or other errors deriving from contradictions between attributes of the name and other data associated with them.
* The Latin form of the name may be ambiguous. Therefore one Latin form may generate more than one form in the original orthography. The confidence of the generated form depends on a number of factors: the likelihood that certain letters or combinations of letters in the language of transliterated represent certain letters in the source language; the differences between different standards of Romanization in different target languages (English, German, French, Spanish etc.), prevalence of the names etc.
* Each source language name components that has been generated from the transliterated form is analyzed in order to extract implicit information such as national origin, gender, religious or sect identity etc.
* The names are then matched with the names in the watch list. The names in the watch list also contain all variants of the name with their corresponding confidence levels. Each variant of each name component is matched with the corresponding name component variants of the name that it is being compared to. Naming conventions regarding name parts that may be omitted are taken into account. Names are assumed to be transcribed with the same transliteration standards. For example, if in one part of a name the letters ch describe sh (e.g. Chehab = Shehab or as in Charlotte, ch will not be used in other parts of the name as a hard ch (e.g. Chuck) or as a guttural (as in German: Achtung, Achmed). This is used in order to find the most likely name variants and to accord them confidence.
* Possible matches are: the same person, a father-son relationship, a son-father relationship, a brother, a cousin or potentially a family member. Each relationship receives a confidence score (e.g. if the watch list contains a name Muhammad Abdallah, even if there is a match of an input name Muhammad Abdallah, the system calculates the statistical probability that this may be the same person, based on the prevalence of the name.

# General Technological Specifications

The system is provided as a stand-alone server application running on Windows Server with a backend and frontend running on a Linux machine (see hardware specifications in annex).

The IntuView technology is fully compatible with a list of database and discovery systems that are in use of the US Government, including most Triple Store databases, SQL databases, Oracle 11G , Endeca and Siebel etc.

The installation of the platform will use an SQL database for matching names.

The platform accepts lists of names in txt/tsv formats or in database. The list may contain additional identifying information about the people they represent (gender, addresses, passport numbers, age, ethnicity etc.). In such a case, the format of the list must be configured by the software provider in advance.

# Components and Processes

The following are the components of the IntuView Name Analyser Name Matcher platform:

* Name Analytics
* Back Transliteration and normalization of transliterated names
* Ethnicity and Gender Matching,
* Name Resolution,
* Matching
* Anomaly Identification,
* Relationship Discovery
* Country Matching

## Name Analytics

This component receives a name of a person in supported scripts, performs analysis of the name, identifies its ethnicity, if it is written in an orthographic system (alphabet) which is not native to its ethnicity, reconstructs the original written form of the name in its original orthography, extracts implicit information from the name such as gender, ethnicity, status, family relationships, religious or sectarian affiliation, tribal affiliation, and in some cases age (in cases of countries which maintain robust databases of name popularity over the years).

The IntuView Name Analyser platform analyses names and ranks the matches of the names and the information extracted from them according to levels of confidence. The confidence of a name analysis result includes a number of elements: the confidence of the back-transliteration (see below); frequency of the name that results from the back- transliteration (e.g. Mahamid) may be مُحَمَّد (Muḥammad) or مَحَامِيد (Maḥāmīd). The former is more common in all the Muslim world however the latter is common in the Western Maghreb countries. The confidence of the former or the latter will derive from the context (the ethnicity of the family name for example).

## Transliteration

Transliteration is the act of mapping of written text (names, words) from one orthographical system (alphabet) into another, with the goal of enabling an individual that does not read the source orthographic system (alphabet) to read it with reasonable accuracy and – if the person is acquainted with the source language – to reconstruct the original spelling in the source orthographical system. Transliteration is not like transcription, which maps the sounds of one language into a writing system through use of a phonetic alphabet (such as IPA) which is not intelligible to the non-linguist.

IntuView Name Analyser Name Analysis receives as input a name written in a supported script and transliterates it into a chosen standard. These standards include:

1. ICS (Intelligence Community Standard) -An Example of ICS transliteration would be the name عبد الحليم أبو عبيدة = ʻAbd-Al-Ḥalīm Abu-ʻUbaydah.
2. Relaxed ICS – based on ICS without the special diacritics that differentiate between various Arabic letters. The name from the previous item would be 'Abd-Al-Halim Abu-’Ubaydah.
3. Google transliteration (not necessarily conforming to pronunciation). The name above would then be transliterated eabd alhalim 'abu eabidatan (without diacritics), however this would not indicate the pronunciation. This is used by Google Translate but does not serve any practical purpose.
4. Strict phonetic transliteration. In this case the name above would be transliterated as ʿbd ạlḥlym ạ̉bw ʿbydẗ. This transliteration is also not very useful as it transliterated only the consonants and ignores the vowels (which in Arabic are optional and in normal writing not included).
5. SATTS (Standard Arabic Technical Transliteration System). SATTS replaces each Arabic letter with a unique symbol based on Latin (الحليم - Al-Ḥalīm would be transliterated AlOlym).
6. BGN – BGN/PCGN romanization refers to the systems for romanization (transliteration into the Latin script) and Roman-script spelling conventions adopted by the United States Board on Geographic Names (BGN) and the Permanent Committee on Geographical Names for British Official Use (PCGN). This standard is not relevant for human names. An example would be Dhū Ḥusein (IC) transliterated into BGN as Dhu Husein.

The scripts supported by the IntuView Name Analyser Name Analysis platform include:

1. Latin, including extended Latin for scripts such as Scandinavian, Czech, Polish, Turkish, Balkan etc. which contain characters not included in standard Latin ASCII.
2. Arabic, including extended Arabic used in Farsi, Urdu, Pashto, Dari and other Arabic alphabet based languages.
3. Hebrew (Israel).
4. Devanagari (Hindi script) used in India.
5. Cyrillic, including extended Cyrillic used in Russian and other languages that use the Russian alphabet.
6. Other scripts that are under development for incorporation into IntuView Name Analyser include: Amharic, Chinese, Japanese (Kanji, Katakana and Hiragana) and Korean (Hangul).

Back Transliteration or Reverse Transliteration is the process by which the name is entered in a script that is not the indigenous script of the name (e.g. a Latin based name in Arabic script or an Arabic name in Latin script). If the name is written in a Romanized form of an Arabic script name (Arabic, Farsi, Urdu etc.), the system will perform back-transliteration to the identified source language and will identify the original components of the name. If the name is a Chinese name represented in one of the two common transliteration standards, the system will identify the standard and create a normalized version (PRC standard Pinyin) that will normalize the name (e.g. Hsi Chin P’ing in Wade Giles standard = Xi Jin-Ping in the Pinyin standard used in the PRC).

The process for Back Transliteration is as follows:

1. The input name is received and its general ethnicity is determined. This is based on statistical models that were generated by training of hundreds of thousands of names from scores of countries and ethnic origins.
2. Based on the ethnicity or ethnicities, the system refers the name to back transliteration to the source orthography of the chosen ethnicity:
   1. Arabic/Farsi/Urdu/Pashto names: Arabic script or extended Arabic;
   2. Western names (and Turkish or Caucasus based: Latin (or extended Latin) script;
   3. Russian names: Cyrillic script.
   4. Hindi names: Devanagari script
   5. Chinese names: Pinyin transliteration standard.
   6. Israeli names: Hebrew script

The back transliteration, unlike transliteration from the source orthography into Latin, is not a one-on-one process. An Arabic name in Latin script may return a number of Arabic names that could correspond to that Latin spelling. These names are influenced by a number of factors:

The transcription or transliteration of a name is frequently corrupted. This corruption can derive from on a number of factors:

1. The native language, accent, literacy and clarity of speech of the conveyor of the name. Can he actually pronounce the phonemes or does he substitute his own language’s phonemes. An Arab may not be able to pronounce the Pashto and Dari پ (P) and will pronounce it (B).
2. Differences between regions. In Afghanistan this would even differ between Eastern and Western Pashto, Dari, Hazari, Uzbek, etc. For example, the written letter ښ is pronounced as a 'sh' in the Western dialect and as a 'kh' in the Eastern dialect. ږ is pronounced as a 'zh' in the Western dialect and as a 'g'' in the Eastern dialect. In Arabic the letter ج will be pronounced and written j in most of the Arabic speaking countries and g (hard g) in Egypt and Oman. ق on the other hand will be alternatively a, q (guttural q), k, g or gh (in Farsi).
3. The musical ear of the transcriber. To what extent he is able to distinguish foreign language phonemes. How well does he understand the culture of the name? Does he identify incongruence in the form he has received and does he distinguish between the name and honorifics?
4. Names that appear to be two separate names may be double-barreled names (comparable to Jean-Luc or Mary-Jane).
5. The name as received may be full or partial, it may be a nickname or nom de guerre. One must understand naming convention in order to extract the real name.
6. In order to apply the correct naming conventions, one must first ascertain the ethnic origin of the name. Double-barreled names with two or more components in the given name or the surname Pakistani/Afghani names Zaaeem, Zuwail, Salak, Siraj, Qaim and more have a second part that can be Al-Islam or Al-din or Muhammad etc. In these cases, the given name is not the first component but both parts (e.g not ‘Abd + Second Name al-Rahman but ‘Abd al-Rahman, not Za’im but Za’im al-Din etc.). Sayid Muhammad as the first two components of a name may a given name Sayid Muhammad (if it is a Pashto name) or Sayid (given name) Muhammad(family name) if it is an Arab name or Sayid (honorific) Muhammad (given name) in a Hazari (Shiite) name.

Each name that is returned receives a confidence score based on the probability that the Latin name may be that Arab (or other) name. The back transliteration engine generates a large number of name alternatives. The final score is based on an algorithm that computes the most likely ethnicities and aggregates similar names.

## Ethnicity and Gender Matching

Ethnicity matching is based on three components:

1. Statistical models of generic ethnicity (~60 models) which identify which ethnicity or ethnicities the structure of a name corresponds to (e.g. Al-Ubaida will be classified as Arab due to the features of the name whereas Watanabe will be classified as Japanese; Mousawi will be classified as Arab whereas Mousavi will be classified as Farsi).
2. A prodigious knowledge base of names from all ethnicities with the corresponding genders, ethnicities and other information.
3. Statistics from country statistics bureaus, phone books and other datasets on the prevalence of given names and family names in different countries.

The IntuView Name Analyser platform first discovers the general ethnicity of the name, then finds the name (if it exists) in the knowledge base. Finally the specific ethnicity is determined based on the statistics of the name parts in different countries of origin.

Identification of gender is based on the same set of algorithms of the ethnicity. For example, some names may be female in one cultural context and male in another and many may be applied to both genders (Carol is female in English and male in Rumanian).

## Name Resolution

Name Resolution is the process by which the different name components are identified, examined and attributed to their possible roles according to the naming conventions of the ethnicity that the name comes from. Hence, for example, a name composed of four parts will receive different interpretations according to its ethnicity:

1. In Arab societies, the structure will be Given Name; Father’s Name (Patronymic); Grandfather’s Name; Family Name
2. A Hispanic name would be Given Name; Middle Name (or a two-part Given Name such as Jose Luis); Family Name; Mother’s Maiden Name.

The algorithms that resolve the name components must also deal with names that are similar in two ethnicities or cultural contexts but are combined differently. For example, Ali alone may be a generic Arab or Muslim name (Sunni or Shiite) however Ali-Reza would be a Persian origin name and mostly Shiite.

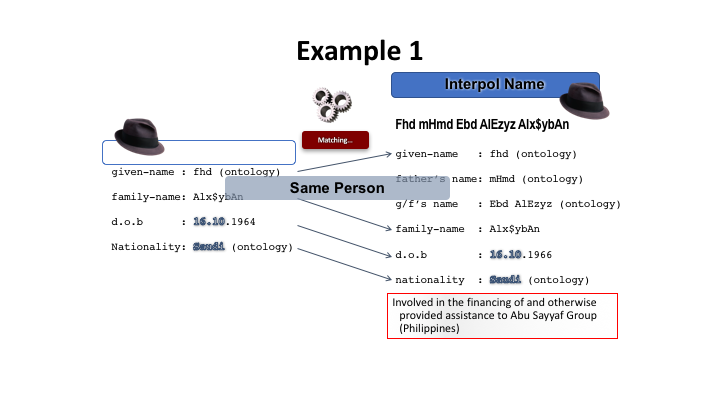
## Name Matching

Name Matching is the process by which one name is matched to another name in spite of discrepancies between them and the fact that they are not a one on one match. The discrepancies can derive from missing parts in one of the names (middle name or patronymic), different spelling of the names, dropping of tribal names that are optional or even components that are perceived by people who are not from the same culture as a name part but are really titles (e.g Sheikh, Mir).

Name Matching is performed by identifying all the possible variants of the input name and the possible roles of each name component. Hence a name such as Abdallah Abu Ahmad may be:

1. Given-Name Abdallah, Kunya (parental name) Abu Ahmad (ie. the father of Ahmad) or
2. Given Name Abdallah and Tribal Name or Family Name Abu Ahmad.

Hence the match of this person could be: Abu Ahmad Abdallah (in the first alternative) - the father of Ahmad whose name is Abdullah or Abdallah Yousif Muhammad Abu Ahmad (in the second alternative). If Abu Ahmad is a clan name and the clan is Bani Sakhr then the person Abdallah Yousif Bani Sakhr could be a match.



## Relationship Discovery

Relationship matching is the process by which the analyzed structure of the names is leveraged to identify whether two people in a database have a potential family relationship. This is particularly relevant in conservative and traditional societies in which each individual carries a name that includes certain elements of his or her lineage. Examples (a non-complete list) of this would be:

1. In Russian: Ivan Vladimirovich Lavrov may be the brother of Irina Valdimirovna Lavrova.
2. In Arab culture: Mounir ‘Abdallah Yousif al-‘Azzam may be the father or son of Abdallah Mounir Yousif al-‘Azzam, the brother of Ahmad ‘Abdallah Yousif al-‘Azzam and the cousin of Yousif Ahmad Yousif al-’Azzam.
3. In Western culture: George Herbert Walker Bush was the son of Prescott Sheldon Bush and Dorothy Walker Bush so he had an arbitrary first middle name (Herbert) and his mother's maiden name as a second middle name (Walker). Prescott Sheldon Bush was the son of Samuel Prescott Bush and Flora Sheldon Bush so he had a first name from his father's middle name (his grandfather's name) and a middle name from his mother's maiden name.
4. In Korean culture cousins of the same family may have the same Generational Name (e.g. Kim Jong Il, Kim Aun Il are the same generation of the Kim family).
5. Eritrean names on the other hand arbitrarily give family names even though naming conventions used in Eritrea and Ethiopia do not have family names and typically consists of an individual personal name and a separate patronymic. This is similar to the Arabic or Icelandic naming conventions. Although traditionally the lineage is traced paternally, legislation has been passed in Eritrea that allows for this to be done on the maternal side as well. In this convention, children are given a name at birth, by which name they will be known. To differentiate from others in the same generation with the same name, their father's first name and sometimes grandfather's first name is added. This may continue ad infinitum. In the West, this is often mistaken for a surname (family name), but unlike European names, different generations do not have the same second or third names. In marriage, unlike in some Western countries, women do not change their name, as the second name is not a surname.

## Anomaly Identification

Anomaly identification is based on a number of factors: internal anomalies within the names and contradictions between the names and the declared information of the asylum seeker. Examples of errors and anomalies include:

* Patronymic anomaly
* Religion ethnicity mismatch
* Missing name component
* Non-letter character
* Acronym
* Offensive words
* Non-compatible ethnicities
* Non-vocalized names in Latin
* Identical patronymic and grandfather's patronymic
* Kunyas
* Orthographic errors
* Klingon names

## Country Matching

The IntuView Name Analyser technology identifies the ethnicity of a name. However, in the regions whence the asylum seekers hail, ethnic groups are mixed together and there is no one on one correspondence between an ethnicity and a country.

For example, a person from Iraq may be (and carry a name) Arab-Sunni, Arab-Shiite, Kurdish, Turkmen, Yazidi. A person from the Persian Gulf States may be Arab, Persian or Pakistani. A Pashtun name has elements that are close to other Persian-based names (Iranian, Urdu etc.) but also unique in tribal features to the Pashtun space in Afghanistan. However, there are Pashtuns in both Pakistan and Afghanistan. Similarly, an Afghan may have a Turkic name (Azeri, Uzbek, Kazakh) or a Persian-based name (Tadjiks, Pashtuns, Hazaris etc.). In fact, there is no such thing as Afghan ethnicity per se. There are also nuances to be dealt with regarding the large population of ethnic Persians in the Arab Gulf States (Bahrain, Kuwait) and Pakistanis in those countries (in Saudi Arabia, there are about 3 million Pakistanis. Many of them are very radicalised such as Syed Rizwan Farook and Tashfeen Malik - both of Pakistani-Saudi origin - who executed the terrorist attack in San Bernardino, California).

The platform contains a sophisticated mapping algorithm that identifies the probability that a person may be from a given country if the ethnicity of his name is identifies. Each ethnicity is mapped to the countries in which it occurs with a factor that indicates the prevalence of that ethnicity in that country (European examples would be Switzerland, Belgium, Russian minorities in the Baltic states, in the Ukraine and the intermingling of ethnicities in the Balkans).

The platform uses an ontology of ethnicities that indicates if a certain ethnicity is a parent or child of another. For example, the Maghreb ethnicity is a child of Arab which is a child (partially) of Muslim. But Maghreb itself includes Tunisian, Moroccan, Algerian, Libyan. Within these there is a high level of similarity in names. So if a person declares himself a Tunisian, and the ethnicity that is returned for his name is Maghreb (one level up from Tunisian that includes Tunisian), this generates a confidence 2 of his declared ethnicity. If the identifies ethnicity is Algerian or Moroccan, the confidence level goes down (3) but remains strong.

## Confidence

IntuScan Name Matcher can be configured to reduce the number of proposed matches to those with highest confidence.

The confidence score is calculated from the following factors:

* Confidence based on the number of matched name components in the matched names - a name with four components (given name, patronymic or middle name, grandfather’s patronymic, family name) can be matched with a name with two components (given name, family name) but that would lower the confidence of the match.
* Confidence based on the commonness of the names – the confidence of a match of a name comprising of two very common components (John + Smith; Muhammad + Abdullah) without additional matching of information will be significantly lower than of names which are not common.
* Confidence based on the transliteration algorithms - name alternatives that reflect standard (non-colloquial) pronunciation receive higher confidence. While such configuration may reduce the level of recall of name matches from non-written sources (names collected by SIGINT or HUMINT and transcribed with phonetic variants), the effect for matching of names on passports or official documents would be less, due to the fact that these names usually follow stricter transliteration standards. For example: the name Ehsen could be the name Ihsan or a phonetic-based transliteration of the North African pronunciation of Hussein or Hassan. If the source of the name is in writing, the latter will receive low confidence.

# Operation

## Configuration

The platform natively accepts lists of names in txt/tsv formats. After installation of the platform, the user may configure the format of the watchlist and the input of lists. The information in the watchlist and the checklists may contain additional identifying information about the people they represent (gender, addresses, passport numbers, age, ethnicity etc.). In such a case, the format of the list must be configured by the software provider in advance. The platform can accept names from a data base. In such a case, the IntuView support team will provide a suitable API.

The options for presentation of results may be configured by the user to determine which are the default components of the result to be displayed in the output. The elements include: Number, Input Name, Caption (generated by IntuScan per name alternative), name components (given name, middle name, patronymic etc.), Ethnicity, Gender, Warnings, Naming Convention etc. This can be later changed as needed. The configuration of the table is possible through click on the name of the user (the login name) on the top right corner, opening Settings/Tables/Name Matcher and checking the desired default elements in the order they should appear in the table from left to right (i.e. checking number, then input name then caption will then present these in this order). The order can be changed by clicking on the table itself after analysis, unclicking all elements and then clicking them in the desired order in the table.

## Loading a watchlist or a number of watchlists

Graphical user interface, text, application, email

Description automatically generated

To upload a watchlist, go to the top menu bar, open the Upload tab. The Upload tab provides a number of options to upload a document or text. These include options for the Document Analysis function and the Name Matcher:

* Upload Text
* Upload file.
* Upload folder
* Upload URL
* **Upload Watchlist – upload a list of names for analysis and alert.**

Choose a list in a format that has been defined. The list must be on the client computer or on a network computer to which the client has access. After choosing the list, it is analyzed as a watchlist.

Go to the Name Matcher – open the list of watchlists and choose the one for matching.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated

## Analyze or Analyze and Match a Name or List of names

To analyze a single name:

* Choose the Single Name tab.

Graphical user interface, text, application

Description automatically generated

* Choose the Watchlist against which the name is to be matched.
* Insert the name
* Click Go

When the analysis is complete, a notification appears. To get the results, click on “Fetch Results”.

Graphical user interface, text, application, chat or text message, email

Description automatically generated

The results will appear in a table (see below Reading the Results). All the possible alternatives of the input name will appear with each one’s possible matches.

To analyze a single name without matching to a watchlist, click on Analyze only. This will then present all the possible name alternatives of the input name without any matching.

To analyze and match a list of names:

* Choose the Lists tab
* Choose the Watchlist against which the name is to be matched (the watchlist must have been uploaded previously).
* Upload a list from the client computer or from a computer that is accessible on the network from the client computer.
* Click Go.

To analyze a list of names without matching to a watchlist, click on Analyze only. This will then present all the possible name alternatives of the input names in the list without any matching.

## Reading the Results

### Report

After the analysis and matching is complete, click on the tab "Report" to receive statistics of the analysis.

Graphical user interface, chart, pie chart

Description automatically generated

The statistics include breakdown of the analyzed list by "Analysis Status" ("analyzed vs. no analysis"), "Gender", "Ethnicities" and "Matching"(only in case of matching against a watchlist - breakdown of types of matches).

In order to view the list of each "slice" or breakdown of the pies - click on the slice. The list of names that correspond to the slice will appear below the pies in the same tab.

Graphical user interface, application, table, Excel

Description automatically generated

The results can be output and downloaded by clicking on the three dots aside the pie - the options are:

* Image: Options- PNG, JPG, SVG, PDF
* Data: Options - JSON, CSV, HTM, XLS,PDF
* Print

### Analysis of Entire Batch

The results of the analysis are presented in a table on the screen. The table is ordered according to the pre-set default columns. The user may view the columns and change them (prevent presentation of an element in the table or change the order) by opening the menu x columns selected and de-selecting undesired elements. In order to change the order of the columns, de-select all and then select them again in the desired order.

Elements of a name that can appear in the analysis include:

* Given Name (includes western given names, Chinese-Korean-Vietnamese “double-barreled” given names)
* Second Given Name (in some cultures – an individual will have more than one given name)
* Third Given Name
* Patronymic (includes Slavic, Greek, Arab patronymics – Slavic in the form of the father’s name with the suffix “ovich”, Greek with the suffix “ou” and Arab with the prefix “ibn/bin” or alone)
* Middle name
* Second Patronymic (Arab grandfather name can come after the father’s name)
* Third Patronymic(Arab great-grandfather name can come after the father’s name)
* Middle  Name (includes middle names, western matronymic middle names – the mother’s maiden name)
* Maiden Name
* Second Middle Name
* Family Name (includes western family names, Chinese/Korean/Vietnamese family names)
* Tribal/Caste Name
* Hispanic maternal name
* Ethnicity
* Naming Convention
* Warning
* Error
* Father
* Son
* Marital Status (of a Woman – sometimes included in the name)

## Exporting Results

There are two options for export (on the right side of the table):

* Export – exports to a csv file the table as it is configured.
* Export All – exports all the data, regardless whether in the viewed table or not.

## Map

The tab "Map" shows a map of the world with the breakdown of the ethnicities in the results of the analysis of the current list being viewed. The breakdown shows the key countries of the ethnicities in bolder colors and other countries where those ethnicities could come from in lighter colors.

## History

The tab "History" allows the user to retrieve the results of previous lists that were analysed. The lists appear in a dropdown menu beneath the History tab. To view the results of a certain list, choose the list from the dropdown menu and click on "Fetch Results". The results will appear in the same format of regular results. To view the report (statistical breakdown) of the historic list, go to the report tab, which will now show the report of the list that is now being viewed.