

The Upcoming International Biometric Vocabulary Standard

Dr. James L. Wayman, Principal UK Expert to ISO/IEC JTC1 SC37 WG1 for the British Standards Institution
Jan. 28, 2012

Some years ago, I was hired by a national government to help them salvage a large, fingerprint-based ID card project that had experienced a “failed bid” when the published Request for Proposal (RFP) was withdrawn over confusion by both government officials and bidders regarding its provisions. I read the withdrawn RFP and was also confused as to what was being requested from industry. For example, the RFP stated that “False Rejection Rate” was to be no more than 2%, but what did this mean? Some government officials interpreted this to mean that no more than 2% of ID card applicants would be wrongly turned away because of a mistaken match against fingerprints already in the database, but others interpreted this to mean that no more than 2% of the people using an issued ID card to access services would be rejected because of failure to match the fingerprints stored on the card. The vendors had asked for clarification on the word “false rejection”. Did the provision require limiting the rate of wrongful matches (false positives) or the rate of wrongful non-matches (false negatives)? And to what ratio did “rate” refer: single attempts with single fingers to all attempts, multiple attempts with multiple fingers to all multiple attempt “sessions”, or users ultimately rejected to all users? The confusion in the RFP was indicative of vocabulary confusion in the entire industry. Biometrics did not have the vocabulary necessary to support large-scale, public procurements.

In 2002, the International Organization for Standardization (ISO) and the International Electrotechnical Commission’s Joint Technical Committee 1 (ISO/IEC JTC1) established Sub-Committee 37 – Biometrics (SC37) to create international standards for the automated recognition of humans. At the first meeting of SC37 in the fall of 2002, the standards development work was divided among six working groups, the first being “WG1- Vocabulary Harmonization”. Over the last eight and a-half years, after 17 week-long international meetings, multiple “ad hoc” sub-committees and dozens of documents submitted by the various national bodies within SC37, we now have definitions for over 100 words in a document expected to become an international standard this year. The chair of WG1 is Steve Clarke of Australia and the editor of the standards document is Peter Waggett of IBM-UK. This group has taken on the task of creating careful definitions to support biometric procurements at all scales, large and small.

The development of a standardized vocabulary for biometrics is inherently a high risk endeavour. According to ISO 860:1996 “Terminology work – Harmonization of concepts and terms”, the chances of successful harmonization of a vocabulary is greater if at least one of the following conditions are met:

- The subject field is well established and relatively stable.
- The subject field deals with concrete objects, such as machinery, tools, materials or industrial products.
- There is a tradition of standardization in the subject field.

But unfortunately for SC37, none of the above conditions holds in the biometrics community. Our technologies and application requirements are rapidly changing; we deal with vague (not concrete) objects, such as security, human recognition and personal “identity”; the development of standards at the international level began only in 2002.

At the beginning of this process, with the help of the French National Body, WG1 developed a corpus of current terms and definitions in use in the literature and the developing Standards. This document immediately demonstrated the problems in the field.

For example, eleven definitions were found (and listed in the "corpus") for the term “template”. These definitions were contradictory and seemed to encompass several different concepts. Other terms, such as “features”, “enrolled features”, and “stored data” were in use for the same concept. In order to *harmonize* these terms - picking one term for each concept and one definition for each term in an internally consistent way - WG1 adopted the following principles:

- To adhere to the ISO vocabulary development processes documented in:
 - ISO 704:2000, Terminology work – Principles and methods
 - ISO 860:1996, Terminology work – Harmonization of concepts and terms
 - ISO 1087-1:2000, Terminology work – Vocabulary – Part 1: Theory and application
 - ISO 10241:1992, International terminology standards – Preparation and layout
- To honour to the greatest extent possible current terms and usages within the biometrics and broader scientific communities
- To respect dictionary definitions whenever possible.
- To recognize that terms will be translated into a wide range of languages.

Following the ISO process requires the development of "concept maps", listing concepts in the field of biometrics and showing graphically their relationships. According to ISO 860, terminology work is based on concepts because:

- Differences between concepts do not necessarily become apparent at the term level.
- Similarity at the term level in different languages does not necessarily mean that the concepts behind the terms are identical.
- Harmonization efforts should, therefore, always start at the concept level and continue at the term level.

So WG1 began by creating a map of the basic sub-fields within the field of “biometrics”.

Currently recognised sub-fields are:

- General concepts
- Devices
- Applications
- Performance
- Biometric systems

- Personnel
- Interactions
- Data in a biometric system

Each sub-field was broken down into further sub-sub-fields and concepts are placed into those sub-sub-fields. As an example, “template” is a concept in the sub-sub-field of “Biometric data” in the “Data in a biometric system” sub-field of the field of “Biometrics”. Once the identified concepts were placed into the appropriate sub-fields and their relationships had been noted graphically, a preliminary name (term) and a list of qualities (properties or qualifiers) was developed for each concept.

Following the above ISO 860 guidelines, the real work was in crafting a consistent and complete list of necessary and sufficient qualities. As an example, based on current common usage, the qualities of “template” might be: “created from processed features; of the same form as processed features; stored for comparison; associated with a biometric data subject”.

When the qualities are complete, we can easily craft those qualities into the definition of the concept name (the term). In the case of “template”, one possible definition from these qualities might be “processed features associated with a biometric data subject and stored for comparison”. This is not inconsistent with the definition given in the Oxford Concise English Dictionary (tenth edition, revised, 2002) as “(2) something that serves as a model or example”.

Often we find that definitions are built on previous concepts already discussed and that the new definitions point to deficiencies in concept maps already developed. Consequently, this process is iterative, often requiring extensive revision of existing terms and definitions as new concept maps, terms, and definitions are developed.

Although most of the work of WG1 is in rectifying conflicting terms and definitions, the exciting part of the work is when we occasionally find concepts for which there is no existing definition or term at all. For example, if a “template” consists of stored processed features, what do we call the biometric data stored in an incompletely processed, but perhaps compressed form, such as the JPEG facial images being stored on e-passports? WG1 decided to consider that data as a “stored, captured sample”.

But there is no term in current literature and work encompassing both “template” and a “stored, captured sample”. WG1 decided to call them both “biometric references”. This adds two new concepts to the “Biometric data” concept map and two new terms to the vocabulary. Adding the term “reference” also resolves issues brought forward by the speaker and facial recognition development communities, which prefer the term “models” over “templates” as a more accurate description of their pattern recognition processes. So “stored captured samples”, “templates” and “models” can all be referred to generally as “biometric references”.

In the WG1 harmonized vocabulary, the following statement has general applicability across most biometric applications -- “Biometric recognition is based on the comparison of one or more captured biometric samples to one or more previously stored references” – whereas a similar statement using “templates” in place of “references” would not be generally true.

So what about “false rejection”? One possible solution would be to use the terms “false acceptance” and “false rejection” only for access control applications, using “false-negative identification error” and “false-positive identification error” for multiple document issuance checks. These latter two words have been used in the international standards on biometric testing, ISO/IEC 19795. The wording of this solution is awkward, but preserves the current primary usages of the terms “false acceptance” and “false rejection”. Another solution would be to always take the perspective of the users (more specifically the “biometric capture subjects”), so that a “false rejection” is being wrongly told “no” because of a technical matching error, regardless of the application. This solution, which is the one taken in the upcoming vocabulary standard, does not require the creation of awkward new words, but leads to use of “false rejection” to sometimes result from a “false match” and sometimes from a “false non-match”, which is confusing to many people in the field. The definition for “biometric false rejection” is “error of rejecting a **biometric claim** that should have been accepted in accordance with a statement on the origin of the **biometric probe** and the **biometric reference**.” Terms in bold face type are defined elsewhere in the vocabulary document. The term “biometric claim” has been defined to allow both positive claims “I/you/she is the bodily source of a specified or unspecified biometric reference” and negative claims “is not the bodily source”. Terms not in bold face type, such as “error” and “rejecting”, are to be understood as according to common English dictionaries. Defining the term “rate” and determining the denominator to be used when determining the “false rejection rate” has been the subject of on-going discussions with the biometric testing community and the Working Group on testing standards, WG5.

Any changes in vocabulary have a potentially huge reverberant effect across all literature in the field. Consequently, none of the definitions developed by WG1 so far are without controversy and all are subject to change in the 5-year ISO/IEC revision process. It is only after this five-year revision process that WG1 will know what parts of the work have been acceptable to the broader biometric community.

So we expect the new international biometrics vocabulary standard, to be known as ISO/IEC 2382 - Part 37, to be released in early 2013 at the end of the routine ISO/IEC acceptance process. But the work will continue. WG1 has already begun updating and adding to the terms in anticipation of the 5-year revision.

About the Author

Dr. Jim Wayman is a research administrator in the Office of Graduate Studies and Research at San Jose State University. He received the Ph.D. degree in engineering in 1980 from the University of California, Santa Barbara, and has worked continuously in the field of automated human recognition since 1984. He joined San Jose State University in 1995 to direct the Biometric Identification Research Program, serving as Director of the U.S. National Biometric Test Center at San Jose State from 1997-2000. He is co-editor of J.Wayman, A. Jain, D. Maltoni and D.Maio (eds) Biometric Systems (Springer, London, 2005), a Fellow of the British Institution of Engineering and Technology, a member of the U.K. National Body to the ISO/IEC JTC1 SC37 standards committee on biometrics, a "core member" of the U.K. Biometrics Working Group and international editor of ISO/IEC 19794-13 standard on voice data format. He was a member of the U.S. National Academies of Science/National Research Council (NRC) Committee "Whither Biometrics?", the NRC Panel on Information Technology, and the NRC committee on "Authentication Technologies and their Implications for Privacy". He holds 4 patents in speech processing and has served as a paid biometrics advisor to 10 national governments.

