日中韓辭典研究所
The CJK Dictionary Institute

Expanded OFAC
VARIANTS OF ARABIC NAMES IN OFAC

The CJK Dictionary Institute, Inc. (CJKI)
Jack Halpern at jack@cjki.org (CEO)

---

1. What is OFAC?

OFAC stands for the Office of Foreign Assets Control, which publishes the Specially Designated Nationals List (SDN), simply referred to as the OFAC list. This list contains individuals and organizations that are a security risk which United States citizens are prohibited from doing business with. It is used worldwide in security and AML applications, border control and watch lists.

CJKI's Database of Arabic Names (DAN) is a constantly expanding comprehensive database of Arabic names and their romanized variants that currently contains 2.4 million entries (see http://www.cjk.org/cjk/arabic/dan.htm). DAN also provides comprehensive coverage of all the Arabic names their aliases found in OFAC, which is updated regularly, and is currently being expanded to cover various other similar lists, such as the Denied Persons List and the United Nations list.

The US government's watch lists have come under fire from members of Congress as being "crippled by technical flaws." Rep. Brad Miller (D-NC), chairman of the Investigations and Oversight Subcommittee of the House Science and Technology Committee, pointed out the railhead system's failure to adequately match variations in the names of suspects, among other problems.

2. Millions of Unlisted Variants

One of the major factors contributing to the system being characterized as *crippled by technical flaws* is its inability to correctly identify and process the numerous variants of the names appearing in the watch lists. That is, there are numerous potential and actual variants of OFAC names which are *not* found the OFAC list. This poses a real danger of failure to identify criminals and suspicious individuals, resulting in false negatives. CJKI maintains a comprehensive database of such OFAC name variants that are *not* listed in OFAC, referred to as **XOFAC** for "Expanded OFAC".

It is important to note that XOFAC is structured differently from DAN. While DAN is a database of name *components* (surnames, given names, name elements), XOFAC consists of variants of *full* names; in other words, names of actual and potential individuals and their variants.

XOFAC Highlights

- Contains millions of potential and actual variants *not* found in OFAC
- Many XOFAC entries are real aliases
- Provides variants of the full name, not name components (as in DAN)
- Highly effective in eliminating false negatives

Let us take the name Hatim Ahmad BARAKAT (خاتم أحمد بركات) as an example. The OFAC list has the following entries for this name:

Main OFAC entry:	**Hatim Ahmad BARAKAT**
OFAC aliases:	Hatam Ahmad BARAKAT
	Hatem Ahmad BARAKAT
	Hattem Ahmad BARAKAT
	Hotem Ahmad BARAKAT

The table below lists the top 15 out of about 130,000 actual and potential variants of the OFAC name **Hatim Ahmad BARAKAT,** which appear in CJKI's XOFAC database. The table has the following fields:

Rank	relative ranking based on component frequencies
Variant	variants of OFAC name, mostly not appearing in the OFAC list
Freq1	frequency of occurrence on the web of **Hatim** variants
Freq2	frequency of occurrence on the web of **Ahmad** variants
Freq3	frequency of occurrence on the web of **Barakat** variants

15 out of 130,000 variants for Hatim Ahmad BARAKAT

Rank	Variant	Freq1	Freq2	Freq3
000001	**Hatem Ahmed Barakat**	0001580000	0039000000	0001180000
000002	**Hatim Ahmed Barakat**	0000925000	0039000000	0001180000
000003	**Hatem Ahmed Bereket**	0001580000	0039000000	0000651000
000004	**Hatem Ahmad Barakat**	0001580000	0025600000	0001180000
000005	**Khadem Ahmed Barakat**	0000194000	0039000000	0001180000
000006	**Hatem Ahmed Bareket**	0001580000	0039000000	0000057200
000007	**Hattem Ahmed Barakat**	0000180000	0039000000	0001180000
000008	**Hatem Ahmed Berekat**	0001580000	0039000000	0000033400
000009	**Hatem Ahmet Barakat**	0001580000	0018400000	0001180000
000010	**Hadim Ahmed Barakat**	0000114000	0039000000	000118000
000011	**Hatem Ahmed Baraket**	0001580000	0039000000	0000016300
000012	**Hatam Ahmed Barakat**	0000081300	0039000000	0001180000
000013	**Hatem Ahmed Barakaat**	0001580000	0039000000	0000014300
000014	**Hatem Achmed Barakat**	0001580000	0000777000	0001180000
000015	**Hetem Ahmed Barakat**	0000065600	0039000000	0001180000

Only the fourth entry in the top 15 above actually appears in the OFAC list. There is no guarantee that the millions of entries in XOFAC actually refer to the individual in question, or that they even exist. That is, they consist of potential and actual combinations of orthographically valid variants, some of which are legitimate and some of which could be used

by other persons of the same name, or not at all. But if the goal is to achieve maximum recall, which it is, comprehensive variant coverage is of great benefit and has no negative effects (except possibly for system resources).

Hattem Ahmed Barakat (Rank 000007 above) is an example of an XOFAC entry that is a legitimate alias not found in the OFAC list that does indeed refer to the person in question. That is, it is an *actual*, not a potential, variant of that person's name. Another example is the name **Mohamed Ben Belgacem AOUADI,** for which OFAC lists one alias, Mohamed Ben Belkacem AOUADI. However, XOFAC lists another three variants for this name, i.e.,

> Mohammed Ben Belgacem AOUADI
> Muhammad Bin Belgacem AWADI
> Mohamed Ben Belgacem AL AOUADI

which are attested as actual, not potential, aliases for this name.

The above demonstrates the important role that the XOFAC database can play in avoiding false negatives; that is, in ensuring that the person in question is correctly identified no matter how the his or her name is spelled or misspelled,

3. Why so many variants?

Why do Arabic names have so many variants? You need not understand this to use XOFAC, but if you are interested in the details we have prepared a document that analyzes the name **Abd Al Aziz AWDA** (عبدالعزيز عودة), which appears in the OFAC list and has over 4000 Arabic and Roman Variants, as summarized in the table below:

Variant Distribution

ITEM	VARIANT TYPE	VARIANT COUNT
1	Arabic variants of last name	4
2	Arabic variants of first name	8
3	Basic Latin variants of last name	18
4	Expanded Latin variants of last name	62
5	Basic Latin variants of first name	72
6	Expanded Latin variants of first name	616
7	Expanded Latin variants of first name (potential)	3561

The full document is available at http://www.cjk.org/cjk/arabic/OFACvars.doc. An academic paper (*The Challenges and Pitfalls of Arabic Romanization and Arabization*) on Arabic orthography Jack Halpern presented at The Second Workshop on Computational Approaches to Arabic Script-based Languages is available here:

> http://www.cjk.org/cjk/arabic/arannana.pdf

For more information on Jack Halpern and The CJK Dictionary INstitute (CJKI), please see:

Jack Halpern: http://www.kanji.org/kanji/jack/index.htm
CJKI: http://www.cjk.org/cjk/index.htm

4. Contact Information

The CJK Dictionary Institute, Inc.
Komine Building (3rd & 4th floors)
 34-14, 2-chome, Tohoku, Niiza-shi
Saitama 352-0001 JAPAN

Phone: 048-473-3508 (+81-48-473-3508)
Fax: 048-486-5032 (+81-48-486-5032)
Email: jack@cjk.org (Jack Halpern)