

What's in a name?

Mar 8th 2007

From The Economist print edition



Johnny Hannah

Computing: Intelligence agencies are using new software to handle the arcane business of comparing lists of names

IN 1990 a Pakistani named Mir Aimal Kansi used an alternative transliteration of his Urdu family name, Kasi, to obtain a visa at the American consulate in Karachi. He entered America, overstayed his one-month visa and then went to the Pakistani embassy in Washington, DC, and obtained a new Pakistani passport, this time with the "n" reinserted in his surname. Using this new identity, he obtained working papers and a driving licence, bought a gun and went on to shoot five CIA employees, killing two, outside the agency's headquarters. (Kansi spent four years on the Federal Bureau of Investigation's Ten Most Wanted list before being captured, and was executed in 2002.)

This case shows how the apparently humdrum process of transliterating names from one language to another can be exploited by criminals. According to the FBI, Kansi also used the names Mir Aimal Kanci, Mir Aman Qazi, Amial Khan and Mohammed Alam Kasi. That last name introduces a further twist: there are more than 15 accepted ways to transliterate "Mohammed" from Arabic into English, and when you count the ways the name is written in the other 160-plus languages that use the Roman alphabet, the figure jumps to more than 200 correct spellings. Transposing words or names from one language or alphabet into another is evidently an inexact science.

In Indonesia, where single names are common, what appears to be just part of a name may in fact be the whole name. Chinese and Korean surnames are often mistakenly written last by Westerners, but some Chinese and Koreans are now adopting the Western convention. And then there is the problem of spelling variants. The Chinese family name Zhou, for example, may be written by English speakers as Jhou, Joe, Chou or Chow. Jafari, the common English transliteration of an Iranian family name, is rendered in German as Djafari or Dschafari. Shahram, the standard English spelling of an Iranian first name, becomes Scharam in German (and Chahram in French).

Such ambiguities cause huge problems for intelligence analysts trying to monitor and prevent terrorist activity. In an effort to avoid being picked out by computer watch-lists, many terrorists use alternative (but linguistically legitimate) transliterations of their names. "It's extremely commonplace, particularly with Islamic names," says Dennis Lormel, former director of the FBI's Terrorist Financing Operations Section, who is now an intelligence consultant at Corporate Risk International, near Washington, DC. "There are just so many variations of a name and they know that, so they can just flip-flop their name around," he says.

But companies in a fast-growing corner of the software industry have developed name-matching programs that can take into account the thousands of possible transliterations of a particular name—say, Mohammed bin Abdul Aziz bin Abdul Rahman Al-Khalifa—as they scan through watch-lists and databases looking for a match. The industry was flooded with investment in 2004 when the 9/11 Commission noted that the terrorists who attacked New York and Washington, DC, on September 11th 2001 defeated watch-lists by using different transliterations of their names. The commission urged the government "to close the long-standing holes in our border security that are caused by the US government's ineffective name-handling software." In-Q-Tel, the investment arm of the Central Intelligence Agency (CIA), began pouring money into name-matching software developers, according to a former official who chose which firms to finance. He says the technology is now becoming "pretty solid, robust stuff".

A name by any other name

"One of our biggest problems has always been variations of names," says Michael Scheuer, who was the head of the CIA's Osama bin Laden Unit from 1996 to 1999. Mr Scheuer says analysis was "backbreaking", especially for Arabic names, because it involved manually compiling lists of variations deemed worthy of tracing. This included positing names with or without titles such as bin ("son of", also written as ben or ibn), abu ("father of", also written as abou), sheikh (tribal leader, also written as sheik, shaikh, shaykh, cheik and cheikh) or haji (Mecca pilgrim, also written as hajj, hajji, hadj, haaji, haajj, haajji and haadj). The article al (also written as el) may be attached to surnames directly, separated from surnames with a hyphen or a space, or omitted altogether. Some variants do not even look similar. Sheikh can be written as jeque in Spanish. Wled, one English transliteration of an Arabic first (and last) name, is often written as Ould in French.

To make matters worse, many bureaucracies tolerate name abbreviations and short forms. The result is that intelligence analysts, no matter how expert, are often plagued by doubts. Has a Russian-speaking intelligence officer in Moscow transliterated into Cyrillic the name of a Nepalese suspect in exactly the same way as a Russian-speaking Uzbek field officer? Has an Italian analyst working with Russian intelligence caught and corrected the error, or passed it along?

Name-matching difficulties actually worsened when counterterrorism activity increased in late 2001. Analysts were granted greater access to databases kept by foreign agencies—but locating relevant files proved hard. A Portuguese case officer, for example, might have difficulty taking advantage of Dutch intelligence on, say, Nepalese Maoist extremists, if he is unfamiliar with Dutch conventions for the transliterations of Nepalese names. The number of people gathering and handling intelligence also increased suddenly, and many newcomers had little language training or were unsure how to transliterate names from spoken sources. Information on suspects increased, but spelling variations—due both to terrorist subterfuge and intelligence shortcomings—made it harder to interpret.

Mr Scheuer says that by late 2004, when he left the CIA, name-matching software was beginning to perform well, and American agencies were investing heavily in the latest technology—with one glaring exception. Computer systems at the State Department, according to Mr Scheuer, were "archaic compared to the rest of the intelligence community". That was a grave weakness, considering that the State Department issues passports and visas for travel to the United States.

Johnny Hannah

If someone fears that the Romanised version of his name has been flagged, he can choose a new (but linguistically correct) transliteration, and then establish that

"Name-matching can play a crucial role in counterterrorism by enabling analysts to piece together snippets of intelligence."

spelling gradually by using it on low-level documents such as a gym membership card or a lease agreement. These “feeder documents” are used to obtain progressively higher-level identity documents, such as a city-issued residence card, a driving licence or a certified birth-certificate translation. These documents, in turn, are presented at consulates to obtain the ultimate prize—passports and visas using the new variation of the name.

“It’s a very tough set of problems,” says Philip Zelikow, executive director of the now-dissolved 9/11 Commission. The group’s research turned up numerous cases of transliteration fraud. Mr Zelikow notes, however, that the American government is now doing a better job handling names. Other experts affirm that the State Department has dramatically upgraded its name-matching software.



There are no firm estimates of how much name-matching software is being sold worldwide. Government agencies generally decline to release figures, and software firms shy from discussing hard numbers. Those in the industry, however, claim that growth is spectacular. Sam Kharoba of First Capital Technologies, based in Baton Rouge, Louisiana, says his firm’s sales have doubled in each of the past three years. Its clients include America’s Defence Department and over 20 other government agencies. Around 25 companies are working in the field in America, and a handful are in Europe.

As watch-lists multiply beyond the realms of intelligence and international travel, demand for such software is likely to grow. Increasingly, watch-lists are used to restrict access to training and education, and to stop people buying property, guns, chemicals and other things that can be made into weapons. Many postal services rely on name-matching software to pick out packages for inspection.

The financial services industry is also adopting the technology, which is often required by central banks and monetary authorities. In America, the Treasury’s Office of Foreign Assets Control is one of the world’s largest users of name-matching technology. It uses it to compile watch-lists that are sent to thousands of banks worldwide. Credit-card companies use the software to spot recidivists applying for new cards under modified names. (Names are cross-referenced with addresses, dates of birth and other data.) Developers and users are hesitant to discuss costs. But OMS Services, a British software firm, says government agencies pay a lot more than commercial users, who pay about \$50,000 for its NameX programme.

Name-matching software is also becoming more sophisticated and performing other functions. The name-matching software made by Identity Systems, based in Old Greenwich, Connecticut, is used by more than 200 government agencies around the world. As well as flagging names on watch-lists, it also sifts historical records to reveal hidden relationships: if two men have entered a country several times on the same plane, sitting apart from each other, might one be a moneyrunner and the other his overseer?

Names and numbers

GNR, a software firm owned by IBM, makes software that “enriches” names by annotating them with inferred cultural information, scored according to probabilities derived from demographic data. Given a particular name it can, for example, say how likely someone is to have a particular place of birth. Names and titles can also provide clues as to birth order, occupation, deaths of spouses and immigration history. GNR also repairs names that are “damaged” by transliteration because the original non-Roman script is lost. The software generates possible original spellings and provides accuracy probabilities for each one. This helps spooks starting with the Romanised versions of, say, Pushtu names, to gather intelligence on those individuals in their native Afghanistan. GNR sells its software to law-enforcement and intelligence agencies—those in Australia, Israel and Singapore are particularly big spenders.

Name-matching software is just one small item in the counterterrorism toolbox. But it can play a crucial role by enabling analysts to piece together snippets of intelligence. What’s in a name? The answer, in some cases, is a surprising amount of valuable information.