



Transforming Potential Into Ability

eFLOW **Approximate Database** *Search Solution*

Approximate Database (DB) Search Engine

eFLOW Approximate Database (DB) Search Engine provides high-performance approximate search technologies for online and offline search and data quality solutions, and for databases with millions of entries.

eFLOW's data capture and delivery platform is unique in its rich offering of over a dozen world-renown recognition engines. The diversity of its engines enable the platform to guarantee a superior automatic recognition rate, further enhanced with the voting technology.

eFLOW is inherently open to additional engines, which continuously enrich the platform yearly.

The latest addition to the *eFLOW* platform is an approximate database search engine.

By using the DB search engine, OCR results such as:

Henry Srnith – Portmoulh Roed – Londor SW15 3TA

can still be validated as:

Henry Smith – Portsmouth Road – London SW15 3TR

As a result, even requests with spelling errors, with incomplete data, or with bad scan results can still be automatically validated against the reference database, avoiding cost-intensive manual correction.

In the same way, the efficiency of the manual correction can be significantly improved by providing the data entry operator with a list of approximate matches from the database.

Unique Features

- Gives quality results extremely fast (milliseconds)
- Queries very large databases (tens of millions of entries)
- Can return reasonable answers to queries even when the query is only approximately equal to the keyword (due to spelling errors, bad scanning, incomplete scan, incorrect database entry)
- Language-independent
- Alias dictionaries allow the transformation of: e.g., München to Munich, Big Apple to New York
- No change to IT infrastructure
- Bad scan and OCR results can be automatically validated against the reference database, avoiding cost-intensive manual correction
- The efficiency of manual correction (completion) can be significantly improved by providing the operator with a list of approximate matches from the database

System Requirements

OS: Windows (2000, 2003, XP), Linux, Solaris, AIX

APIs: Sockets, C++, COM, PHP, Java, Tcl

CPU: Multi-CPU ready

RAM: Approx. 800 MB per 10 million data

Best results searching for:
Street: ??einrichstha?evS?? Post Code: ?9???? City: ?es?he?e

Street	Post Code	City	Quality
Heinrichsthaler Str.	59872	Meschede Heinrichsthal	243
Breslauer Str.	59872	Meschede	226
Brückenstr.	59872	Meschede	222

Best results searching for:
Street: ?fe?str. Post Code: ????? City: T?aleiSCMW?II?r

Street	Post Code	City	Quality
Uferstr.	66967	Thaleischweiler-Fröschen	235
Uferstr.	26639	Wiesmoor	217
Uferstr.	06538	Weischlitz	217

The BKK Group Success Story



The BKK Group processes 40,000 variably structured documents, matched against customer databases of millions of entries, improving recognition by over 40%.

The BKK Group is one of the three largest public-health insurance companies in Germany. The central IT center of the BKK Group handles the medical certificates, prescriptions, receipts, and other forms – between 30,000 and 40,000 per day.

eFLOW recognizes and classifies the address, insurance number, date of invoice and birth date from 200 health insurers, and from numerous doctors.

In order to ensure superior recognition rates, TIS employed its high-performance approximate matching engine. If several fields (e.g., name, street, city) contain errors or are incomplete, the engine finds the closest candidate in the database. Even with 25 million entries, the matching is accomplished at the speed of 10-15 matches per second. If the approximate candidate does not meet the stringent quality criteria, the record and the one or two most likely candidates are saved for operator validation. The operator then simply confirms most of the suggestions.

Savings: Using an approximate database matching engine has improved *eFLOW*'s recognition by over 40%.