



” AMRON - baza danych, w której liczy się jakość

Karol Kacprzak

Kierownik Projektu AMRON III,

Specjalista ds. Analiz i Rozwoju Systemu AMRON

ANALIZA CENTRUM AMRON

9 marca 2026





W przypadku baz danych wykorzystywanych do analiz rynku nieruchomości kluczowe znaczenie ma nie tylko liczba zgromadzonych rekordów, ale przede wszystkim ich wiarygodność i spójność. W AMRON od początku przyjęto zasadę, że wartość bazy buduje jakość danych, a nie ich liczba. Dlatego rozwój systemu oraz procesy związane z gromadzeniem danych są projektowane w taki sposób, aby maksymalnie ograniczać ryzyko pojawiania się błędów i niespójności.

Od początku istnienia systemu przyjęto założenie, że jako administrator bierzemy pełną odpowiedzialność za jakość danych znajdujących się w bazie, niezależnie od sposobu ich pozyskania. Oznacza to, że każdy rekord wprowadzony do systemu podlega wieloetapowej kontroli jakości. Dane nie trafiają do bazy w sposób automatyczny i niekontrolowany – są weryfikowane zarówno w momencie wprowadzania, jak i w kolejnych etapach funkcjonowania w systemie.

PO PIERWSZE: WALIDACJE

Już na etapie wprowadzania danych stosowany jest rozbudowany mechanizm walidacji, który ma na celu eliminowanie błędów. W nowej wersji systemu AMRON mechanizmy te zostały znacząco rozbudowane i podzielone na trzy poziomy.

Pierwszym z nich jest walidacja wstępna, wykonywana podczas wgrzywania do systemu pliku pakietowego z danymi. Na tym etapie sprawdzana jest poprawność struktury pliku oraz jego format. System weryfikuje m.in. zgodność układu pól oraz dopuszczalnej liczby rekordów w jednym pliku. Celem tej walidacji jest wychwycenie błędów technicznych jeszcze przed rozpoczęciem właściwego procesu przetwarzania danych oraz zapewnienie stabilności i wydajności całego rozwiązania. Walidacja ta dotyczy wyłącznie wprowadzania pakietowego – w przypadku pojedynczego wprowadzania danych przez interfejs WWW lub przez API nie jest ona wykonywana.

Drugim poziomem są walidacje I stopnia, które obejmują wszystkie rekordy wprowadzane do systemu – zarówno pojedynczo przez użytkownika w przeglądarce, jak i poprzez API czy w sposób pakietowy. Na tym etapie system sprawdza podstawową poprawność danych. Weryfikowana jest m.in. obligatoryjność pól, zgodność typów danych, dopuszczalna liczba znaków oraz formaty poszczególnych informacji (np. maski dla określonych atrybutów). W przypadku rekordów wprowadzanych pakietowo lub przez API dodatkowo sprawdzana jest poprawność wartości słownikowych – system weryfikuje, czy dana wartość należy do zestawu dopuszczalnych wartości zdefiniowanych w słownikach systemowych.

Najbardziej zaawansowanym poziomem kontroli są walidacje II stopnia. Ich zadaniem jest wychwytywanie sytuacji, w których dane formalnie spełniają wszystkie podstawowe wymagania, ale mogą wskazywać na potencjalny błąd merytoryczny. Dotyczy to na przykład przypadków, gdy informacje w różnych polach wzajemnie się wykluczają lub gdy wartości znacząco odbiegają od typowych parametrów dla danego typu nieruchomości. W takich sytuacjach użytkownik wprowadzający dane musi potwierdzić ich poprawność. Niektóre z tych walidacji mogą również skutkować przekazaniem rekordu do dodatkowego zatwierdzenia przez użytkownika posiadającego najwyższe uprawnienia w systemie (Administrators Systemu Centralnego). W ramach tego poziomu działają także mechanizmy porównujące nowy rekord z rekordami już istniejącymi w bazie, co pozwala na wykrywanie i eliminowanie potencjalnych dubli.



PO DRUGIE: PRZEGLĄDY JAKOŚCI

Kontrola jakości danych nie kończy się jednak na etapie walidacji systemowych. Każdy rekord może zostać objęty procesem przeglądu jakościowego, a w przypadku pojawienia się jakichkolwiek wątpliwości zakładane jest tzw. zgłoszenie weryfikacyjne. Jest to formalny proces mający na celu sprawdzenie poprawności danych i – jeśli zajdzie taka potrzeba – ich poprawę. Ponadto, banki wprowadzające dane do systemu również przeprowadzają cykliczne weryfikacje jakości wprowadzonych przez nich danych. Dzięki temu baza jest na bieżąco monitorowana i ulepszana, a potencjalne nieścisłości są systematycznie eliminowane.

PO TRZECIE: STANDARYZACJA

Zakres danych obligatoryjnych w bazie został zaprojektowany w sposób zgodny z Rekomendacją J, co zapewnia spójność informacji i ich przydatność w analizach rynku nieruchomości. W nowej wersji systemu wprowadzono także szereg rozwiązań wspierających użytkowników w procesie wprowadzania danych. Jednym z nich jest integracja z zewnętrznymi rejestrami i bazami danych. Przykładowo atrybut „obręb ewidencyjny” jest słownikiem zgodnym z danymi GUGiK, a po wybraniu adresu lista dostępnych obrębów jest automatycznie zawężana do wybranej gminy. Wprowadzany numer budynku jest weryfikowany w oparciu o Bazę Punktów Adresowych, a kody pocztowe są dopasowywane do wskazanej lokalizacji. W przypadku podania numeru księgi wieczystej system automatycznie wskazuje właściwy sąd wieczystoksięgowy. Kolejnym jest fakt, że rekordy mogą być automatycznie uzupełniane o informacje (tj. w zakresie roku budowy, liczby kondygnacji w budynku, konstrukcji budynku, zagęszczenia zabudowy, dostępności komunikacyjnej, otoczenia, informacji o efektywności energetycznej i inne) zgromadzone w Bazie Danych Budynków – autorskiej bazy informacji o budynkach w Polsce prowadzonej przez Centrum AMRON. Automatyczne pobieranie danych z Bazy Danych Budynków redukuje błędy ludzkie i poprawia wiarygodność rekordów, a ponadto użytkownicy nie muszą ręcznie wyszukiwać i uzupełniać wielu parametrów budynku. System robi to automatycznie, co znacznie skraca czas wprowadzania rekordu. Dodatkowo kursy walut mogą być pobierane automatycznie w zależności od wybranej waluty i daty transakcji.

System korzysta również z oficjalnego rejestru GUS TERYT. Podczas wprowadzania danych kod TERYT jest uzupełniany automatycznie na podstawie rejestrów TERC, SIMC i ULIC, dzięki czemu użytkownik nie musi ręcznie wprowadzać pełnej ścieżki adresowej.

POZYSKIWANIE DANYCH Z RCN A ICH JAKOŚĆ

W ostatnim czasie szczególnego znaczenia nabrała także kwestia pozyskiwania danych z Rejestrów Cen Nieruchomości (RCN). Po udostępnieniu nowych form dostępu do tych danych pojawiły się możliwości pobierania dużych pakietów informacji w formacie GML, obejmujących transakcje z wielu powiatów. W jednym z testów tego rozwiązania pobraliśmy dane obejmujące początkowo ponad 170 powiatów. Na pierwszy rzut oka mogłoby się wydawać, że takie pakiety stanowią idealne źródło danych umożliwiające szybkie zwiększenie skali bazy. W praktyce jednak kluczowa okazała się analiza jakości tych informacji. Po przekształceniu przykładowych plików GML do postaci tabelarycznej (akceptowanej przez System AMRON) okazało się, że wiele rekordów nie zawiera podstawowych informacji – na przykład numeru adresowego lokalu. Dodatkowo w plikach GML brakowało również informacji o przeznaczeniu gruntu, które są dostępne w danych pozyskiwanych klasyczną drogą z RCN.

Jeszcze większe różnice ujawniły się podczas szczegółowej analizy jakości danych. W jednym z powiatów pakiet zawierał ponad 28 tysięcy transakcji. Po zastosowaniu podstawowych filtrów jakościowych, takich jak kompletność udziału w nieruchomości, minimalna powierzchnia gruntu, realistyczna cena za metr kwadratowy czy wy-



kluczenie niestandardowych źródeł informacji, liczba rekordów możliwych do wykorzystania spadła do około 5 tysięcy. Oznacza to, że blisko 80% danych zostało odrzuconych z powodu niewystarczającej jakości lub niekompletności.

Ten przykład bardzo dobrze pokazuje, dlaczego w AMRON konsekwentnie stosowana jest zasada „jakość ponad ilość”. Choć technicznie możliwe byłoby szybkie zwiększenie liczby rekordów w bazie poprzez automatyczne wczytywanie dużych pakietów danych, w praktyce oznaczałoby to wprowadzenie znacznej liczby informacji niepełnych lub wątpliwych. Zamiast tego stosowane są procedury selekcji danych, które pozwalają pozostawić w bazie wyłącznie rekordy spełniające określone standardy jakości.

Dlatego w przypadku danych z RCN w wielu sytuacjach nadal wykorzystywana jest klasyczna ścieżka pozyskiwania informacji, mimo że jest ona bardziej czasochłonna i często objęta dodatkowymi ograniczeniami, np. dotyczącymi liczby rekordów możliwych do jednorazowego pobrania. Pozwala ona jednak uzyskać bardziej kompletne dane, zawierające m.in. pełne informacje adresowe oraz dodatkowe atrybuty istotne z punktu widzenia analiz rynku.

PODSUMOWANIE

Wszystkie opisane mechanizmy – od wielopoziomowych walidacji, przez integrację z rejestrami publicznymi, aż po procesy weryfikacyjne – mają jeden wspólny cel: zapewnienie najwyższej możliwej jakości danych w bazie AMRON. Dzięki temu użytkownicy Systemu mogą opierać swoje analizy na wiarygodnych, spójnych i dokładnie zweryfikowanych informacjach. Naszym celem w AMRON jest tworzenie bazy opierającej się na wiarygodnych danych – większa liczba rekordów ma znaczenie tylko wtedy, gdy są one jakościowe.



KAROL KACPRZAK

Kierownik Projektu AMRON III, Specjalista ds. Analiz i Rozwoju Systemu AMRON

e-mail: karol.kacprzak@amron.pl

Centrum AMRON

Centrum AMRON to profesjonalny ośrodek analityczno-badawczy, świadczący usługi w zakresie monitorowania i analiz rynku obrotu nieruchomości. Od ponad 20 lat z powodzeniem dostarczamy naszym klientom i partnerom zweryfikowaną informację o rynku nieruchomości, wyjaśniamy zjawiska na nim zachodzące, a także zapewniamy dostęp do rzetelnej informacji gospodarczej. Specjalizujemy się w badaniu zmian cen transakcyjnych nieruchomości i ich wpływu na perspektywę rozwoju polskiego rynku mieszkaniowego – jako jedyny podmiot na rynku obrazujemy zmiany na rynku kredytów hipotecyjnych, dzięki uzupełnieniu danych z AMRON danymi z Systemu SARFiN. W odróżnieniu od innych podmiotów zajmujących się analizą rynku nieruchomości, nasze opracowania prezentują rzeczywisty obraz rynku, oparty na cenach transakcyjnych. Zapewniamy wielostronny obraz rynku nieruchomości oparty na kompleksowej analizie czynników makro- i mikroekonomicznych oraz szerokim horyzoncie czasowym, uwzględniającym dane historyczne, bieżące oraz prognozy.

NOTA PRAWNA

Dane wykorzystane w niniejszej analizie zebrane zostały z dołożeniem należytej staranności i rzetelności. Mimo to Centrum AMRON zastrzega, iż przedstawione informacje mają charakter wyłącznie poglądowy i nie mogą być traktowane jako usługi doradztwa ani jakiegokolwiek inne usługi. Centrum AMRON nie ponosi jakiegokolwiek odpowiedzialności za skutki wykorzystania informacji zawartych w analizie, a w szczególności z tytułu jakiegokolwiek decyzji lub działania podjętego na podstawie tych informacji. Niniejsza publikacja jest chroniona prawami autorskimi – każdorazowe upowszechnienie całości lub części analizy wymaga powołania się na źródło prezentowanych danych.

Centrum AMRON działa w imieniu i na rzecz Związku Banków Polskich, właściciela Systemu AMRON.