

Aribert Zbigniew Szarejko
H+H Software GmbH
Getynga, Niemcy
Zbigniew.Szarejko@hh-software.com



HAN – Likwidujemy bariery dostępu do źródeł elektronicznych, zachowajmy kontrolę!

Streszczenie: Artykuł przedstawia możliwości, które przynosi stosowanie w bibliotekach oprogramowania HAN (Hidden Automatic Navigator) firmy H+H Software GmbH. Zadaniem HAN-a jest możliwie maksymalne obniżenie bariery dostępu czytelników do elektronicznych źródeł danych, przy równoczesnym zachowaniu kontroli przez bibliotekę (uprawnienia czytelników, ich uwierzytelnienie, kontrola licencji, statystyka użytkowania). W tym celu HAN integruje się zarówno z różnorodnymi systemami bibliotecznymi, jak i z innymi elementami infrastruktury bibliotecznej (np. multiwyszukiwarki naukowe, menedżery bibliografii, portale wydawnictw).

Słowa kluczowe: dostęp do źródeł elektronicznych, statystyki użytkowania źródeł elektronicznych, zdalny dostęp do źródeł elektronicznych, HAN, integracja elementów infrastruktury bibliotecznej.

Problemy związane z dostępem do elektronicznych źródeł informacji

Rozwój technologii informatycznych, a przede wszystkim technologii związanych z internetem oraz sprzętem, którym posługuje się zarówno czytelnik, jak i biblioteka, zmieniły diametralnie w ciągu ostatnich kilku dziesięcioleci sposób udostępniania i korzystania ze zbiorów bibliotecznych. Czasy, kiedy zasoby były dla czytelnika dostępne tylko w godzinach pracy biblioteki, czasy kiedy te zasoby istniały tylko w formie drukowanej, czasy kiedy dostęp do nich był możliwy tylko w samej bibliotece – należą definitywnie do przeszłości. Obecnie postulat dostępności zasobów o każdej porze dnia i nocy, i to z każdego miejsca na świecie, wydaje się zupełnie oczywisty. To powoduje, że infrastruktura biblioteczna, a nawet organizacja pracy biblioteki, są w ciągłym pościgu za rozwojem technologii oraz za zmieniającymi się przyzwyczajeniami czytelników. Pojawiają się i rozwijają nowe systemy biblioteczne, nowe sposoby korzystania z informacji, nowy sprzęt, którym posługują się czytelnicy.

Biblioteki stają przed wyzwaniem, które można określić jako trudne do rozwiązania sprzeczności. Z jednej strony są czytelnicy, którzy pragną szybkiego i nieskomplikowanego dostępu do zasobów swojej biblioteki przez 24 godziny i 7 dni w tygodniu, w tym także wtedy, kiedy znajdują się poza biblioteką. Z drugiej strony jest biblioteka, która dąży do udostępniania swoich zasobów w taki właśnie sposób, jednakże przy zachowaniu maksymalnej kontroli. Konieczność tej kontroli wynika już choćby z faktu, że dostęp do zasobów elektronicznych biblioteki kupują lub otrzymują za pośrednictwem odpowiedniej instytucji (w Polsce głównie ICM UW¹) i pierwotnie, zgodnie z warunkami licencji, dostęp taki jest ograniczony tylko do komputerów znajdujących się na terenie uczel-

¹ *Wirtualna Biblioteka Nauki* [online]. ICM UW [Dostęp 26.02.2020]. Dostępny w: <http://wbn.icm.edu.pl/licencje/>.

ni. Kontrola nad użytkowaniem źródeł elektronicznych może wynikać także z umów licencyjnych, które biblioteki podpisują z wydawnictwami (np. w zakresie liczby jednoczesnych użytkowników danego źródła lub jego dostępności tylko dla wybranych grup czytelników). W przypadku gdy wydawnictwo zabezpiecza dostęp do źródła indywidualnym kontem, biblioteka także staje przed trudnym do rozwiązania dylematem: hasło nie może być podane użytkownikowi w sposób jawny i przeważnie bibliotekarz musi je osobiście wpisywać czytelnikowi. To z kolei oznacza, że czytelnik korzysta z zasobu w bibliotece, niemal pod czujnym okiem bibliotekarza. Ukoronowaniem tej obszernej listy problemów może być coraz bardziej powszechne stosowanie w różnego rodzaju portalach/stronach internetowych własnych profili czytelnika, do których czytelnik musi logować się za pomocą innych danych logowania niż te, których używa na co dzień w bibliotece lub uczelni. Takie bariery zniechęcają. Powodują, że biblioteka postrzegana jest jako mało przyjazna.

Sytuacja dla biblioteki nie jest więc łatwa: należy sprząc ze sobą wiele różnorodnych elementów infrastruktury bibliotecznej w taki sposób, aby uniknąć wszystkich powyżej wymienionych problemów. Niektóre rozwiązania likwidują pewne problemy, generując jednak nowe niedogodności. Jako przykład mogą tu posłużyć dwa względnie często stosowane sposoby zdalnego (w rozumieniu spoza kampusu uczelni) dostępu do zasobów elektronicznych. Jeden z nich to zastosowanie tzw. serwera proxy; czytelnik jest tu jednak zmuszony do odpowiedniego skonfigurowania przeglądarek internetowych, z których korzysta. Zdalny dostęp jest także możliwy dzięki zastosowaniu technologii VPN (virtual private network), ale i w tym przypadku czytelnik musi przeważnie zainstalować na swoim komputerze tzw. program kliencki VPN.

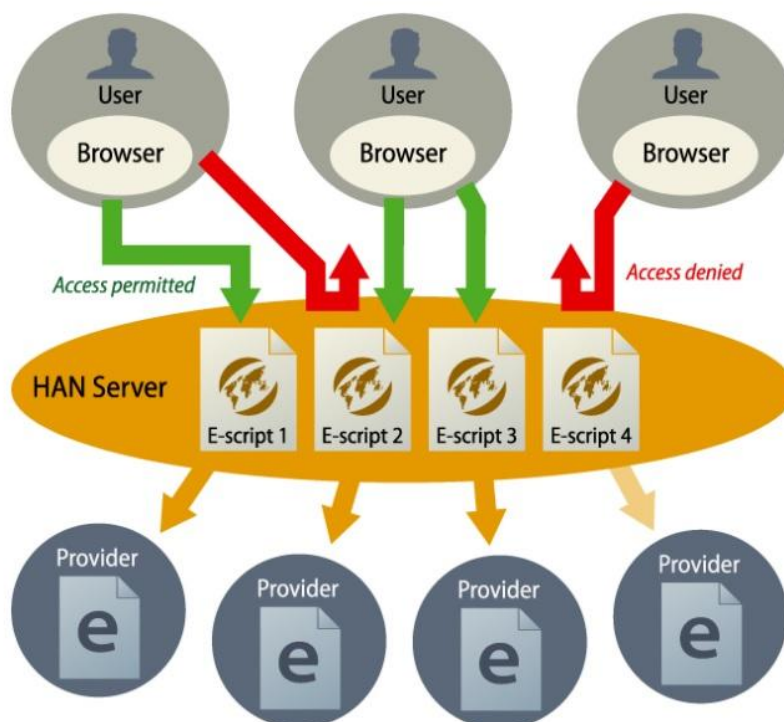
HAN – rozwiązanie problemów związanych z dostępem do elektronicznych źródeł informacji

Oprogramowanie **HAN (Hidden Automatic Navigator)** niemieckiej firmy software'owej **H+H Software GmbH** próbuje rozwiązać opisane powyżej problemy. Z technicznego punktu widzenia HAN jest serwerem proxy specjalnego rodzaju. Jest to tzw. *reverse proxy*², który działa w taki sposób, że po stronie czytelnika nie ma żadnej konieczności instalowania jakichkolwiek programów klienckich czy też pluginów ani nie jest wymagane przekonfigurowywanie przeglądarek internetowych. Czytelnik klika w przeglądarce po prostu na linki dostępu poza siecią uczelnianą, które oferuje mu biblioteka. W momencie kliknięcia link zostaje przekierowany na serwer HAN-a, po czym HAN wykonuje szereg czynności kontrolnych.

Pierwsza kontrola to uwierzytelnienie czytelnika. Biblioteka musi mieć przecież pewność, że do zasobów dostaną się tylko uprawnieni użytkownicy. Dla czytelników znajdujących się na terenie biblioteki/uczelni wystarczająca jest kontrola adresu IP, z którego nadeszło wywołanie. Jeśli wywołanie nadchodzi z jakiegoś „obcego” adresu IP, HAN wysyła do przeglądarki czytelnika stronę logowania, gdzie czytelnik powinien nanieść dane swojego konta. HAN jest w stanie sprawdzić takie dane w różnych systemach bibliotecznych (w Polsce są to: Aleph, Horizon/Symphony, VTLs, ProLib, Patron, SOWA, MOL Optivum). Istnieje także możliwość uwierzytelnienia użytkownika poprzez dowolną bazę danych SQL (interfejs ODBC) lub poprzez bazę danych LDAP. Wszystkie te usługi

² *Reverse proxy*, czyli odwrotny serwer pośredniczący to rodzaj serwera proxy, który pobiera zasoby w imieniu klienta z jednego lub większej liczby serwerów. Te zasoby są następnie zwracane do klienta, wyglądając tak, jakby pochodziły z samego serwera proxy [przyp. red.].

uwierzytelniania można w HAN-ie ze sobą łączyć kaskadowo, tzn. jeśli HAN nie znajdzie użytkownika w pierwszej bazie danych, to może go szukać w następnych. Przy czym czytelnik wcale nie musi wiedzieć, w której bazie danych odbędzie się kontrola jego danych logowania.



Rys. 1. HAN obsługuje uwierzytelnianie poszczególnych użytkowników i umożliwia administratorowi ograniczenie lub zezwolenie na dostęp do zasobów na podstawie uprawnień użytkownika. Można to zrobić na dwóch poziomach: authentication=uwierzytelnienie (określenie, którzy użytkownicy mogą uzyskać dostęp do HAN) i authorization=autoryzacja=upoważnienie (zdefiniowanie zasobów HAN, do których dany użytkownik może mieć dostęp).
Źródło: Hidden Automatic Navigator. Version 3.2 [online]. [Dostęp 01.03.2020]. Dostępny w: <https://download5.hhs-softwares.com/download/HAN/HAN32-Manual.pdf>.

Jeśli uwierzytelnienie czytelnika zakończy się sukcesem, to w następnym kroku HAN wyśle do przeglądarki odpowiednie źródło elektroniczne. Jeszcze w tym momencie dostęp do takiego źródła może być przez HAN-a zablokowany według wskazań konfiguracyjnych biblioteki. Może tak się zdarzyć, jeśli w HAN-ie zdefiniowano, że z pewnych zasobów nie mogą korzystać wszyscy czytelnicy (np. w obrębie uczelni wszyscy korzystają bez ograniczeń, a poza jej obszarem – tylko pracownicy naukowci). Podobną kontrolę HAN może wykonać także w celu zabezpieczenia umów licencyjnych, np. w przypadkach kiedy wydawnictwo sprzedało bibliotece dostęp do źródła elektronicznego tylko dla określonej liczby równoczesnych sesji. W obu takich przypadkach HAN wyśle do przeglądarki czytelnika stronę informującą o braku uprawnień lub czasowym wyczerpaniu limitu jednoczesnych użytkowników. Równocześnie taka informacja zapisana zostanie w odpowiednim protokole w celu późniejszego wykorzystania jej do obliczeń statystycznych czy analiz jakościowych prowadzonych w bibliotekach.

HAN jest także w stanie rozwiązać powyższy problem dostępu do zasobu zabezpieczonego indywidualnym hasłem. Dla dowolnego linku, który „przechodzi” przez serwer HAN-a może być zdefiniowany tzw. skrypt, który w tle wysyła do danego wydawnictwa odpowiedni zestaw dodatkowych linków. Linki te mogą oczywiście zawierać sekwencję adresów URL konieczną do zalogowania czytelnika w danym wydawnictwie. Sam czytelnik tego procesu nie widzi, natomiast po jego zakończeniu otrzymuje w przeglądarce już zalogowaną stronę zasobu. Dane logowania nadal pozostają tajemnicą biblioteki. Dodatkowym bonusem tego mechanizmu jest to, że jeśli czytelnik po zalogowaniu miałby dostęp np. do strony z ustawieniami abonamentu tegoż zasobu, to za pomocą HAN-a dostęp do takich stron może być zabroniony.

W ostatnim czasie biblioteki coraz częściej borykają się z problemem automatycznego wyłączenia przez wydawnictwa dostępu do źródeł elektronicznych. Dzieje się tak wtedy, gdy algorytmy posiadowane na serwerach wydawcy stwierdzają, że w bardzo krótkim czasie ściągana jest nazbyt duża liczba bajtów, co z kolei może oznaczać, że ktoś lub tzw. robot próbuje zapisać na swoim komputerze niedozwoloną liczbę książek lub artykułów. W takiej sytuacji wydawcy nie zawsze powiadamiają bibliotekę o aktywizacji blokady, czasami bibliotekę informują o tym dopiero czytelnicy, którzy nagle pozbawieni są dostępu do danego zasobu. Jest to na pewno sytuacja dla bibliotek i czytelników bardzo niekomfortowa. HAN umożliwia odrzucenie prób pobierania nieograniczonej liczby bajtów w krótkim czasie już po stronie biblioteki, może też zablokować na pewien czas czytelnika, który tego dokona.

W przypadku niektórych portali internetowych HAN znacznie ułatwia pracę czytelnikowi, udostępniając system jednorazowego logowania, czyli tzw. SingleSignOn. Oznacza to, że opisane powyżej uwierzytelnienie poprzez HAN-a jest wystarczające, aby czytelnika automatycznie zalogować, np. do jego półki, jak to jest w przypadku IBUK-a (PWN) i NASBI (Helion) lub do wyszukiwarki naukowej (Primo/Exlibris). Dzięki temu czytelnik nie musi pamiętać danych dwóch różnych kont. Do obsługi zasobów oferowanych przez bibliotekę wystarcza tylko jedno konto, najczęściej powiązane z kontem w systemie bibliotecznym.

Inne funkcjonalności HAN-a

W ostatnich latach rozwój HAN-a zmierza coraz bardziej w kierunku stworzenia centralnego narzędzia łączącego ze sobą różne elementy infrastruktury bibliotecznej z portalami/stronami wydawnictw i innego rodzaju firm oferujących usługi dla bibliotek. HAN dysponuje interfejsem programistycznym (HANWebAPI), z którego takie firmy mogą korzystać. HANWebAPI stosowane jest obecnie dla następujących rozwiązań:

- integracja z multiwyszukiwarkami naukowymi (*discovery services*) Primo-SFX/ExLibris, EDS/Ebsco, Integro/Sygnity,
- integracja z menedżerem bibliografii Citavi/Academic Swiss Software GmbH,
- integracja z portalem czasopism naukowych Browzine/ThirdIron,
- integracja z systemem linkout dla pełnych tekstów Lean Library/Lean Library a Sage Publishing Company.

We wszystkich tych przypadkach mechanizm działania HAN-a jest podobny: przeszukiwanie zasobów jest otwarte dla każdego czytelnika, HAN przejmuje kontrolę dopiero nad linkiem prowadzącym do zasobu pełnotekstowego.

Dla bibliotek ważny jest oczywiście nie tylko ułatwiony dostęp do zasobów elektronicznych oraz jego kontrola, ale także protokołowanie użytkowania zasobów elektronicznych oraz bazujące na tym obliczenia statystyczne. Statystykę użytkowania biblioteki otrzymują przeważnie od samych wydawnictw, przy czym problematyczna jest porównywalność takich statystyk. Natomiast HAN protokołuje wywołania wszystkich „idących” przez niego linków w sposób jednolity, czyli statystyki HAN-a odnoszące się do różnych wydawnictw są porównywalne. Bardzo istotną cechą statystyki HAN-a jest także fakt, że kryteria jej obliczeń definiowane są przez lokalnego administratora HAN-a (bibliotekarza), a nie przez dostawcę zasobów.

Protokołowane są następujące informacje:

- nazwa zasobu (np. bazy danych, czasopisma, dokumentu internetowego),
- nazwa użytkownika (jeśli uwierzytelnienia dokonano na podstawie adresu IP, nazwy takiej brak i w tym miejscu HAN protokołuje adres IP czytelnika lub nazwę, którą w HAN-ie przydzielono temu adresowi),
- adres IP, z którego nadeszło wywołanie,
- czas pierwszego i ostatniego kliknięcia w zasobie (godzina:minuta:sekunda),
- wynikający z powyższego czas użytkowania,
- liczba bajtów ściągniętych przez czytelnika.

Opcjonalnie HAN może zapisywać także inne wartości, które następnie wykorzystuje się w statystyce. Są to przede wszystkim:

- liczba wywołań oraz liczba ściągniętych bajtów dla linków o określonym charakterze (stanowi to podstawę późniejszych obliczeń np. liczby ściągniętych plików PDF; mechanizm ten można także stosować, gdy chcemy pozyskać bardziej szczegółowe dane, np. ile razy w ramach portalu książek elektronicznych została pobrana konkretna publikacja),
- rodzaj czytelnika (np.: gość, pracownik, student),
- przynależność czytelnika np. do jednostki organizacyjnej danej instytucji lub dyscypliny naukowej (dane uzależnione są od rodzaju informacji zawartych w koncie czytelnika).

Dzięki temu istnieje możliwość wygenerowania bardzo prostych w obsłudze statystyk użytkowania dla różnych jednostek organizacyjnych/dyscyplin naukowych lub rodzajów czytelników. Takie statystyki mogą być dobrą podstawą do optymalizacji wewnątrzuczelnianego podziału kosztów zakupu dostępu do źródeł elektronicznych.

Przy obliczaniu statystyki użytkowania można stosować w HAN-ie następujące kryteria:

- podstawa obliczania tabel (według nazw zasobów, według rodzaju czytelnika, według adresów IP, według przynależności czytelnika do jednostki organizacyjnej czy dyscypliny naukowej),
- wybór zasobów, użytkowników, adresów IP, dla których obliczenie ma być wykonane,
- możliwość grupowania nazw zasobów, czytelników lub adresów IP, dla których obliczenia wykonywane są sumująco (np. biblioteka główna, filie, akademiki itp.),

- obliczenia dla tzw. atrybutów; obecnie są używane przede wszystkim dwa z nich: „brak licencji/przekroczony limit jednoczesnych użytkowników” oraz „brak uprawnień” (w ten sposób biblioteka może ustalić liczbę nieudanych prób wywołania zasobów ograniczonych licencjami równoczesnego użytkownika lub uprawnieniami,
- zawężenie obliczeń tylko do tych sesji przeglądarek, które minimalnie lub maksymalnie ściągnęły określoną liczbę bajtów; pozwala to np. na szybkie wyszukanie tych sesji, w których liczba ściągniętych bajtów była niepokojąco wysoka,
- wszystkie te obliczenia mogą być wykonane w podziale na odpowiednie okresy: rok, półrocze, kwartał, miesiąc, bez podziału – od dnia do dnia.

Bardzo oryginalną możliwością obliczeń statystycznych w HAN-ie jest tabela równoczesnego użytkownika. Daje ona odpowiedź na pytanie, ile sesji przeglądarek znajdowało się równocześnie w tym samym zasobie. Informacja ta jest bardzo istotna przy określaniu, czy liczba zakupionych licencji dostępu do jakiegoś zasobu jest wystarczająca, a także – co może być jeszcze bardziej istotne – czy liczba takich licencji nie jest nadmierna.

Wszelkie dane osobowe, które HAN protokołuje, a następnie wykorzystuje w statystyce mogą być anonimowane lub pseudoanonimizowane zgodnie z wytycznymi RODO.

Obecnie w Polsce istnieje prawie 50 instalacji HAN-a w różnego rodzaju bibliotekach uczelnianych, w tym: dziesięć w uniwersytetach, dziewięć w uczelniach technicznych, osiem w uczelniach niepublicznych, po sześć w uczelniach medycznych i w państwowych uczelniach zawodowych, pięć w uczelniach przyrodniczych, trzy w uczelniach ekonomicznych oraz jedna w uczelni papieskiej. Poza Polską HAN bywa instalowany także w instytucjach administracji państwowej (np. ministerstwa, sądy) lub w zakładach przemysłowych. W niedalekiej przyszłości planuje się także instalacje HAN-a na serwerach wydawnictw gazet codziennych lub tygodników (tu dla ułatwienia kontroli sprzedawanych przez nie licencji).

HAN jest oprogramowaniem bazującym na windowsowej wersji serwera webowego Apache oraz na nieSQL-owej bazie danych MongoDB. Może być instalowany na serwerach MS Windows od wersji 2012. Instalacje takie odbywają się ostatnio przeważnie w środowisku maszyn wirtualnych.

Więcej informacji na temat HAN-a można otrzymać w języku niemieckim lub angielskim:

- na stronach <https://www.hh-han.com/>, <https://www.hh-han.com/en/default.cfm>,
- kontaktując się bezpośrednio z firmą H+H Software GmbH <https://www.hh-han.com/en/contact.cfm>,
- kontaktując się z autorem artykułu.