

Nieskończoność we wnioskowaniu o danych i wiedzy

Popularnonaukowe streszczenie projektu

Popularnym przekonanie głosi, że informatyka, inaczej niż matematyka, zajmuje się skończonymi obiektami: zarówno modelowana rzeczywistość, jak i dostępne zasoby pamięciowe są skończone, więc nie ma ani potrzeby ani miejsca na nieskończoność. Tymczasem, owszem, nieskończoność pojawia się w informatyce, i to zarówno jako wyzwanie, jak i jako środek.

Obiekty rozważane w informatyce są faktycznie zwykle skończone, ale mogą być dowolnie duże. Już dane wejściowe, choć skończone, mogą być dowolnie duże. Zatem, na przykład, algorytm ewaluacji zapytań zawsze eksploruje skończoną instancję bazy danych, ale musi działać poprawnie na każdej z nieskończonej liczby takich instancji. Dla tradycyjnych systemów komputerowych, nieograniczone wejście jest podstawowym założeniem. Jednak w ostatnich latach pojawił się trend opierania systemów optymalizowanych do przetwarzania danych w oparciu o tzw. FPGA, które są bardzo szybkimi komponentami sprzętowymi przetwarzającymi wejście o określonym, relatywnie niewielkim rozmiarze. Użycie ich do przetwarzania danych nieograniczonego rozmiaru jest trudnym wyzwaniem, z którym zamierzamy się zmierzyć. Będziemy badali możliwości wzbogacenia FPGA w zewnętrzną pamięć bez pogarszania wydajności.

Trudniejsze zadania obliczeniowe polegają nie na eksploracji danego obiektu, ale na rozstrzygnięciu, czy istnieje obiekt o zadanych własnościach. Wiele problemów w dziedzinie baz danych ma ten charakter: Czy istnieje instancja bazy danych, w której zapytanie wybierze choć jeden rekord? Czy dwa zapytania wybierają te same rekordy we wszystkich instancjach bazy danych? Jako że rozważane obiekty mogą być dowolnie duże, pytania takie wydają się wymagać przeszukania nieskończonej przestrzeni kandydatów. To nie tylko wygląda na trudne, ale czasami można udowodnić, że jest to nie możliwe. Na przykład wiadomo, że żaden program komputerowy nie jest w stanie odpowiedzieć na pierwsze z powyższych pytań dla każdego zapytania w języku SQL. Zamierzamy poszukiwać ograniczonych przypadków, w których ten problem da się rozwiązać. Przy czym celem naszym będzie nie wskazanie poszczególnych własności, ale projektowanie formalizmów, w których można wyrażać wiele własności przydatnych we współczesnych zastosowaniach baz danych do danych grafowych (jak sieci społeczne lub sieci połączeń lotniczych) oraz drzewiastych (jak pliki w formacie XML lub JSON, stosowane między innymi w komunikacji między systemami komputerowymi).

Względnie nowym składnikiem systemów informatycznych są bazy wiedzy. Jak wiadomo, w bazie danych przechowujemy skończony obiekt i odpowiadamy na zapytania na jego temat. Tymczasem w bazie wiedzy przechowujemy skończony zbiór własności obiektu, i próbujemy wnioskować z nich inne własności. Nowa własność może być wywnioskowana, o ile nie istnieje obiekt spełniający przechowywane własności, ale nie spełniający nowej własności. W związku z tym, że obiekty nie muszą być nigdzie przechowywane fizycznie, bazy wiedzy można użyć do wnioskowania o obiektach nieskończonych. Tak więc może się zdarzyć, że szukamy nieskończonego obiektu o zadanych własnościach. Zaskakująco, czasami jest to łatwiejsze niż szukanie skończonego obiektu o tych samych własnościach, dzięki większej swobodzie w budowaniu nieskończonych obiektów. Planujemy wykorzystać to zjawisko, jak również badać kiedy i jak znalezione nieskończone obiekty można przerobić na skończone, co bywa łatwiejsze niż szukanie skończonych obiektów bezpośrednio. Skupimy się na własnościach wyrażanych w tzw. logikach deskrypcyjnych, stanowiących sedno rekomendowanych przez W3C standardów OWL 1 i OWL 2.