

Recenzja rozprawy doktorskiej

Marcin Krystek

zatytułowanej:

Data Mesh for Smart Cities

1. Problem badawczy i jego znaczenie

Podjęmowany problem naukowy dotyczy analizy wieloaspektowych relacji między technologią, zarządzaniem danymi, rozwojem miast, a zastosowaniami biznesowymi, ze szczególnym uwzględnieniem koncepcji Data Mesh. Należy zauważyć, że dotychczasowe działania miast związane z wdrażaniem koncepcji „Smart City” skupiały się na głównie uruchamianiu systemów informatycznych przeznaczonych do konkretnych zastosowań. Systemy te działają jednak niezależnie od siebie, co najczęściej skutkuje chaosem informacyjnym i organizacyjnym, wysokimi kosztami operacyjnymi oraz błędami w realizacji kluczowych procesów biznesowych, w tym w podejmowaniu kluczowych decyzji. Zgodnie z aktualnymi badaniami naukowymi, opracowanie elastycznego podejścia do zarządzania dużymi zbiorami różnorodnych danych w środowisku inteligentnego miasta umożliwia opracowanie efektywnych metod zarządzania danymi z uwzględnieniem współczesnych wyzwań społeczno-organizacyjnych i technicznych, które często są pomijane przez tradycyjne podejścia do projektowania platform danych. **Zatem stwierdzam, że analizowany w recenzowanej rozprawie problem ma charakter naukowy, jest aktualny i ważny, a jego oryginalne rozwiązanie w recenzowanej rozprawie polega na opracowaniu Platformy Otwartych Danych dla Inteligentnego Miasta (ODSC) z wykorzystaniem koncepcji Data Mesh.**

Biorąc pod uwagę znaczenie praktyczne badań prowadzonych w ramach recenzowanej rozprawy doktorskiej można stwierdzić, że realizacja koncepcji Poznania jako miasta inteligentnego musi opierać się na rozwoju infrastruktury technologicznej oraz strategicznym wykorzystaniu wysokiej jakości danych. W konsekwencji może to prowadzić do usprawnienia procesów zarządzania miastem, poprawy jakości usług publicznych oraz tworzenia środowiska opartego na partycypacji, w którym mieszkańcy aktywnie angażują się w kształtowanie przyszłości swojego miasta. Dostępność i odpowiednia jakość danych stanowią bowiem podstawę do podejmowania świadomych i skutecznych decyzji, efektywnego wykorzystania zasobów oraz ciągłych innowacji. **Zatem, ze względu na wdrożeniowy charakter recenzowanej rozprawy doktorskiej, wysoko oceniam opracowanie koncepcji i realizację procesu wdrożenia opracowanej Platformy Otwartych Danych dla Inteligentnego Miasta w Poznaniu.**

W mojej ocenie obszar teoretyczny obejmujący problematykę zarządzania dużymi zbiorami danych zasobów miejskich, jak i obszar aplikacyjny dotyczący usprawniania procesu wdrażania platform zarządzania danymi w praktyce zarządzania inteligentnym miastem dobrze pozycjonują recenzowaną rozprawę w dyscyplinie informatyka techniczna i telekomunikacja. Należy jednocześnie zauważyć, że powiązanie obszarów tematycznych z zakresu informatyki oraz zarządzania jest trudne i wymaga podejścia interdyscyplinarnego. Autor wykazał się dobrym rozeznanieniem i szeroką znajomością omawianych zagadnień.

2. Wkład autora

Za najważniejszy wkład autora uznaję, wskazane w rozprawie doktorskiej, opracowanie Platformy Otwartych Danych dla Inteligentnego Miasta (ODSC), która łączy innowacyjne podejście do tworzenia platform danych oparte na Data Mesh z najnowszymi osiągnięciami technologicznymi w dziedzinie sztucznej inteligencji i dużych modeli językowych. Opracowane podejście prezentuje innowacyjny sposób zarządzania i wykorzystywania złożonych systemów danych, co jest szczególnie istotne w kontekście rozwijającej się infrastruktury cyfrowej w środowiskach miejskich.

W podrozdziale 4.3. zatytułowanym „Basic tenets of Data Mesh in the context of Smart City” autor opracował podstawowe założenia Data Mesh w kontekście Smart City. Autor słusznie stwierdza, że zaproponowana zmiana paradygmatu, a także nowe koncepcje wprowadzone przez Data Mesh (takie jak domena danych, rozproszone zarządzanie danymi), mogą prowadzić do rozwiązania najważniejszych słabości platform danych pierwszej generacji (hurtowni danych) i drugiej generacji (jezior danych). Zmiana paradygmatu nie polega jedynie na przeniesieniu uwagi z technologii na dane. Wprowadza ona spektrum zmian na różnych poziomach, począwszy od perspektywy socjologicznej, obejmującej interakcje międzyludzkie, poprzez aspekty techniczne, takie jak sposób wykorzystania konkretnych narzędzi, sposób, w jaki infrastruktura dostarcza moc obliczeniową i pamięć masową, na zarządzaniu danymi kończąc. Wysoko oceniam zaproponowaną przez autora w podrozdziale 4.4 zatytułowanym “Lessons learned from implementing Data Mesh in a Smart City context” strategię wdrożenia ODSC składającą się z następujących elementów: współpraca, podejście „małych kroków”, zrozumienie, wartość danych, promocja i upowszechnianie, scenariusz działań, oraz zarządzanie zadaniami. Strategia ta umożliwi realizację procesu wdrożenia ODSC w sposób kompleksowy z uwzględnieniem zarówno aspektów organizacyjno-społecznych, jak i technologicznych (przeanalizowanych szczegółowo w podrozdziale 4.5 zatytułowanym „Technical aspects”).

Istotnym wkładem jest również opracowanie metody automatycznego generowania metadanych z wykorzystaniem modeli LLM, którą autor przedstawił w rozdziale 5 zatytułowanym “Large Language Models in the service of Open Data platform for Smart City”. Autor podzielił zakres zadań realizowanych przez model LLM w ramach proponowanych eksperymentów na dwie główne grupy: identyfikację pojęć – polegającą na nazywaniu atrybutów na podstawie ich wartości; oraz generowanie treści – polegające na generowaniu opisów na podstawie dostarczonego kontekstu. Kontekst obejmował zarówno metadane opisujące zbiór danych, jak i wskazówki zawarte w instrukcjach modelu. Zadanie identyfikacji pojęć, które obejmowało również poszukiwanie relacji między atrybutami, było skuteczniej realizowane przez model o szerszym kontekście i trenowany na nowszych danych. W przypadku generowania opisowej treści na podstawie dostarczonych informacji, model o mniejszym kontekście osiągał dobre wyniki. W konsekwencji badań autor zaproponował wybranie odpowiedniego modelu LLM w oparciu o zadanie, które ma on wykonać. Wykorzystując wygenerowane informacje, model LLM może skutecznie realizować zadania wspomagania decyzji, na przykład, w zakresie identyfikowania zbiorów danych najbardziej odpowiednich dla ogólnie sformułowanych kryteriów wyszukiwania. W konsekwencji może to skrócić czas potrzebny na wdrożenie nowego produktu dla użytkownika końcowego.

W kontekście wkładu autora wysoko oceniam rozważania zawarte w rozdziale 6 zatytułowanym „ODSC Deployment”. Autor przedstawił rozwiązania koncepcyjne, technologiczne i aplikacyjne, które opracował i które zostały wykorzystane podczas wdrażania platformy ODSC w mieście Poznań. Autor przedstawił koncepcję wdrożenia oraz kluczowe technologie wykorzystane do wdrożenia ODSC w Poznaniu, funkcjonalności wdrożone w pierwszej fazie prac wdrożeniowych, ze szczególnym uwzględnieniem katalogowania zbiorów danych. Scharakteryzowane zostało również oprogramowanie zapewniające funkcjonalności analizy danych oraz funkcjonowanie i analiza wyników prototypów narzędzi opracowanych w celu automatyzacji procesów generowania

wartościowych opisów zbiorów danych. Wdrożenie opracowanej w niniejszej rozprawie platformy koncepcji ODSC pozwoli, w konsekwencji, systematycznie optymalizować procesy związane z udostępnianiem, zarządzaniem i przetwarzaniem danych. We wstępnej fazie wdrożenia stwierdzono, że wykorzystanie modeli LLM do generowania metadanych rozwiązuje problemy związane z poprawną interpretacją danych, które nie posiadają odpowiedniej dokumentacji i wiedzy dziedzinowej. Ponadto, zdolność modelu do reprezentowania szerszego kontekstu pozwala na generowanie dokładnych i bogatych w treść charakterystyk zbiorów danych. Oba te aspekty są kluczowe dla zapewnienia właściwej widoczności tych zasobów na platformie danych, umożliwiając ich ponowne wykorzystanie w przyszłości.

Stwierdzam, że wkład autora umożliwił oryginalne rozwiązanie problemu naukowego polegające na opracowaniu Platformy Otwartych Danych dla Inteligentnego Miasta (ODSC) z wykorzystaniem Data Mesh.

3. Poprawność

Zdefiniowane przez doktoranta cele i założenia rozprawy doktorskiej zostały osiągnięte w sposób zgodny z kanonem procesu poznania naukowego. Autor rozprawy opanował umiejętność posługiwania się metodami wnioskowania naukowego, jak też zdobył niezbędny warsztat badawczy.

Za najbardziej wartościowe aspekty uważam opracowanie Platformy Otwartych Danych dla Inteligentnego Miasta z wykorzystaniem koncepcji Data Mesh, zaplanowanie i przeprowadzenie eksperymentów badawczych dotyczących metody automatycznego generowania metadanych z wykorzystaniem modeli LLM oraz opracowanie strategii wdrożenia ODSC oraz analizę wniosków z wstępnego etapu wdrożenia platformy w Poznaniu. Sformułowane przeze mnie uwagi krytyczne i pytania problemowe nie pomniejszają mojej bardzo pozytywnej oceny recenzowanej rozprawy doktorskiej i mają głównie charakter polemiczny, redakcyjny lub prowokujący do dalszych prac badawczych. Recenzowana rozprawa w mojej ocenie jednoznacznie posiada walory poznawcze, metodyczne i aplikacyjne. Autor wykazał się dobrą znajomością metodyki badań ilościowych i jakościowych. Praca napisana jest poprawnym językiem, a styl prowadzonego wywodu odpowiada kanonowi opracowań naukowych. W ujęciu formalnym całość prezentuje się dobrze, spełniając wymogi przewidziane dla prac awansowych na stopień naukowy doktora. Pozytywnie oceniam wcześniejsze publikacje i prezentacje na konferencjach naukowych, wybranych badań przedstawionych w recenzowanej rozprawie doktorskiej. Umożliwia to bowiem konsultowanie prowadzonych badań naukowych w szerszym środowisku naukowym, co przyczynia się do wzrostu jakości badań prezentowanych w rozprawie doktorskiej. Autor nie ustrzegł się jednak pewnych nieścisłości i błędów, które wskazałem w dalszej części niniejszego punktu recenzji.

Cel prowadzonych badań określany jest niejednolicie w różnych częściach rozprawy.

Przykładowo, na stronie 9 autor pisze: „Celem badań jest zatem zbadanie obszarów, w których wykorzystanie sztucznej inteligencji (AI), a w szczególności modeli językowych (LLM), co może wnieść znaczną wartość dodaną w rozwiązywaniu złożonych i wysoce praktycznych problemów związanych z tworzeniem i utrzymaniem platformy danych dla miasta”. Natomiast w „Conclusions” jest napisane: „Celem badań było opracowanie koncepcji platformy danych dla nowoczesnego miasta aspirującego do miana Smart City”. Główny cel badań powinien być prezentowany w sposób jednolity w całej rozprawie doktorskiej. Na stronie 9 jest zdanie “In recent years, we have witnessed the rapid development of technologies in the fields of machine learning and artificial intelligence”. Uczenie maszynowe jest jednym z obszarów sztucznej inteligencji, dlatego uważam, że powinno być napisane, na przykład “...artificial intelligence including machine learning”.

Również na stronie 9 autor napisał: „We are aware of the current state of knowledge and recommendations presented in publications such as [7], [8] regarding how to create urban data platforms.” W pracach naukowych należy unikać stwierdzeń, w których autor ocenia samego siebie.

Strona 11, podrozdział 1.5. zatytułowany „Research Framework”. Autor wskazuje wymagania w zakresie, między innymi, koncepcji, projektowania, analizy ryzyka, technologii dotyczące opracowania Platformy Otwartych Danych dla Inteligentnego Miasta. Autor wskazał, że wykorzystuje metodykę badawczą inżynierii systemowej, co uważam za właściwy wybór. Jednakże w niewielkim stopniu wskazuje, w jaki sposób odnosi tę metodykę do badań prowadzonych w ramach przedmiotowej rozprawy doktorskiej.

Przykładowo, autor powinien wskazać, które części rozprawy dotyczą projektu koncepcyjnego, analizy wymagań, projektowania, itd. W związku z tym formułuję pytanie do doktoranta:

Jakie są poszczególne etapy metodyki badawczej zastosowanej w ramach rozprawy doktorskiej i jakie metody badawcze zastosowano w poszczególnych etapach, zarówno w odniesieniu do badań teoretyczno-poznawczych, jak i badań empirycznych?

Na stronie 14, w akapicie: “In a data-centric view, a Smart City is defined by its capacity to harness diverse data types that describe and characterize ongoing processes, transforming this data into actionable knowledge for future governance or new business opportunities. Supporting this process presents considerable challenges, requiring continuous innovation, the resolution of emerging obstacles, and the adoption of cutting-edge technological advancements.” brakuje odwołania do literatury przedmiotu.

W rozdziale 2 autor mógł w szerszym zakresie przeanalizować istniejące koncepcje i rozwiązania Smart City odwołując się w szerszym zakresie do literatury przedmiotu.

Sekcja „Conclusions” - w większym stopniu należało odnieść się do wyników i wniosków, a w mniejszym opisywać, co zrobiono w poszczególnych rozdziałach.

4. Wiedza kandydata

W rozdziale 2 zatytułowanym "Smart City" autor przeprowadza analizę, dyskusję i ocenę istniejących definicji Smart City. Autor słusznie zwraca uwagę, że zdolność do efektywnego gromadzenia i dystrybucji danych jest cechą charakterystyczną nowoczesnego, cyfrowego inteligentnego miasta.

Rozważania prowadzone w podrozdziałach 2.2 zatytułowanym „Poznań Smart City strategy” dotyczące rozważań odnośnie ogólnych zasad i przyjętej strategii wdrażania koncepcji inteligentnego miasta oraz w podrozdziale 2.3 zatytułowanym „Poznań Smart City data assets” dotyczącym analizy danych, które będą przetwarzane na platformie, są bardzo istotne z punktu widzenia osiągnięcia celów recenzowanej rozprawy doktorskiej, jednakże zabrakło mi szerszego odniesienia do innych miast.

Stąd też formułuję pytanie do doktoranta:

Jakie są podobieństwa i różnice w realizacji koncepcji Smart City przez Poznań w odniesieniu do innych, przykładowych miast w Polsce i za granicą, w których ta koncepcja jest wdrażana?

Bardzo wysoko natomiast oceniam podrozdział 2.4 zatytułowany “Challenges in implementing Open Data platform for Smart City”. W odniesieniu do literatury przedmiotu autor analizuje wyzwania techniczne i organizacyjne, dotyczące opracowania platformy ODSC, wynikające przede wszystkim ze złożoności, niejednorodności i dużej dynamiki danych pochodzących z wielu, często heterogenicznych źródeł, z których każde wykorzystuje inne protokoły, formaty i technologie. Na podstawie przeprowadzonej analizy literatury przedmiotu autor słusznie stwierdza, że ta różnorodność potęguje się, gdy platforma musi obsługiwać wielu użytkowników, z których każdy ma własne strategie zarządzania i wymagania. W rezultacie zarządzanie na dużą skalę i harmonizacja danych w ramach ODSC pozostają poważnym wyzwaniem, mającym wpływ na wydajność i skalowalność.

Z kolei w rozdziale 3 zatytułowanym „Data Mesh” autor słusznie zauważa, że w literaturze zaproponowano różne podejścia do budowy platform danych. Większość rozwiązań jest zoptymalizowana pod kątem wydajności, zbliżając istniejące technologie do fizycznych ograniczeń narzucanych przez sprzęt. Istnieją jednak poważne wątpliwości, czy te platformy i architektury rzeczywiście umożliwiają transformację cyfrową i przejście na model zarządzania decyzjami oparty na danych. Dlatego autor proponuje nową koncepcję budowy platform danych wykorzystującą podejście Data Mesh, której celem jest rozwiązanie problemów społeczno-technicznych i organizacyjnych, które są zazwyczaj pomijane w podejściach prezentowanych w literaturze przedmiotu.

W podrozdziale 3.1 zatytułowanym „Data platform architectures” autor przedstawia i analizuje, istniejące platformy/podejścia zarządzania danymi, takie jak hurtownie danych, jeziora danych czy też multimodalna architektura chmurowa. Na uwagę zasługuje wskazanie przez autora zalet i wad tych rozwiązań. Bardzo wysoko oceniam podrozdział 3.2 zatytułowany „Design pattern weaknesses”, w którym autor, na podstawie analizy literatury przedmiotu, słusznie zauważa, że centralizacja zarządzania danymi wprowadza istotne wąskie gardło, utrudniając efektywną integrację i wykorzystanie różnych źródeł danych. Autor wskazuje również inne ograniczenia istniejących platform i podejść do zarządzania danymi. W konsekwencji autor słusznie uzasadnia potrzebę wykorzystania podejścia Data Mesh, które w sposób wnikliwy i dokładny analizuje, na podstawie literatury przedmiotu, w podrozdziałach: 3.3 zatytułowanym „Data Mesh principles” oraz 3.4 zatytułowanym „Data Mesh approach for data platform design”. W rozdziałach 2 i 3 rozważania dotyczą przede wszystkim obszaru analizy danych w zakresie dyscypliny Informatyka techniczna i telekomunikacja.

W rozdziale 4 zatytułowanym „Smart City data landscape using Data Mesh” autor przedstawił proces wdrażania koncepcji ODSC w Poznaniu, a także przyjęte założenia i zdobyte doświadczenia. W kontekście analizy istniejącego stanu wiedzy wysoko oceniam podrozdział 4.1 zatytułowany „Urban data platforms in literature”. Autor dokonał w nim analizy, na podstawie literatury przedmiotu, istniejących rozwiązań w zakresie zarządzania danymi w miastach. Autor słusznie zauważa, że prosta adaptacja rozwiązań technologicznych opracowanych przez korporacje do ich własnych celów jest sama w sobie ryzykowna i bez odpowiednio zreorganizowanych procesów może nie przynieść oczekiwanych rezultatów. W przypadku inwestycji komunalnych, gdzie zaangażowane są środki publiczne, należy zminimalizować ryzyko związane z nieskutecznymi wdrożeniami, aby uniknąć oskarżeń o niewłaściwe zarządzanie. W podrozdziale 4.2 zatytułowanym „Smart City Data Platform pillars” autor analizuje takie filary platform danych, jak: technologia, użyteczność, zrównoważony rozwój i zarządzanie. O ile analiza tych filarów jest niezbędna z punktu widzenia badań realizowanych w niniejszej pracy, zabrakło mi jednak szerszego odwołania się do literatury przedmiotu. W rozdziale 4 rozważania dotyczą przede wszystkim obszaru inżynierii systemów w zakresie dyscypliny Informatyka techniczna i telekomunikacja.

Zestawienie bibliograficzne zawiera głównie pozycje publikowane w czasopismach i wydawnictwach o zasięgu międzynarodowym. Tematyka cytowanych prac związana jest ściśle z obszarami badawczymi podjętymi w rozprawie doktorskiej. Pozycje literatury są aktualne. Autor mógł uwzględnić większą liczbę pozycji literaturowych, zwłaszcza dotyczących problematyki Smart City. Jednakże ze względu na wdrożeniowy charakter niniejszej rozprawy, w którym, oprócz aspektów naukowych, duży nacisk kładzie się na aspekty wdrożeniowe, uważam, że liczba i zakres pozycji literatury jest wystarczający w kontekście analizy istniejącego stanu wiedzy. Analiza i interpretacja źródeł literaturowych jest prawidłowa. Autor w trakcie rozważań w rozprawie dokonuje ich analizy i syntezy oraz przedstawia własne wnioski i krytyczną ocenę analizowanych pozycji literatury.

Podsumowując, moja ocena znajomości, doboru, analizy i interpretacji źródeł literaturowych przez autora jest pozytywna. Na podstawie rozważań teoretyczno-poznawczych prowadzonych przede wszystkim w rozdziałach 2, 3 i 4 stwierdzam, że **rozprawa doktorska prezentuje ogólną wiedzę**

teoretyczną kandydata do stopnia doktora w dyscyplinie Informatyka techniczna i telekomunikacja.

5. Inne uwagi¹

Zawartość merytoryczna rozprawy jest spójna z metodyką badawczą inżynierii systemów pozwalającą na realizację założeń i celów badawczych. Autor sposób sprawny przedstawia cały przeprowadzony cykl badawczy, jak również przejrzysto prezentuje wnioski z badań. Zatem stwierdzam, że oceniana rozprawa doktorska prezentuje umiejętność samodzielnego prowadzenia pracy naukowej przez kandydata do stopnia doktora.

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami)² moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak X)

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Marek Kerner

Podpis

¹ Opcjonalnie

² <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>