

# CIDR: Chat-oriented Innovations in Database Research

Eugene Wu  
Columbia University  
ewu@cs.columbia.edu

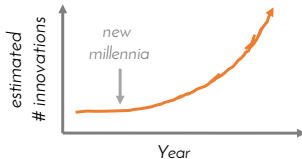


Figure 1: Estimated innovations per year.

The number of large innovations per year has been increasing rapidly since the start of the new millennia (Figure ??). From Hadoop and big data, to internet of things, to millennial startups, to visualization, to deep learning, each new innovation has left an undeniable mark on society, and ultimately, the world. Although the database community may not have led the charge in these innovations, we have been successful at staking our claim to the infrastructure and core tenants in each of these innovations. But must we continue to play a reactive role? Should we not be the drivers in new innovative areas? The answer is clearly a resounding *yes*, but towards what nascent area? Well, do I have a bridge to sell you, and it is chat bots.

Yes. Chat bots.

Conversational communication is slated to be the dominant communication channel. Chat-based applications e.g., Snapchat<sup>1</sup>, WeChat<sup>2</sup>, and Line messenger<sup>3</sup> are respectively worth \$20B, \$83B, and \$9B. In other words, each of these applications are *individually* valued on the order of the *entire* \$33B database market<sup>4</sup>.

In the long term, the largest companies in the world (e.g., Google, Microsoft, Facebook, Tencent) are throwing their weight behind chat. This makes sense because chat is ubiquitous—texting is chat, IMs are chat, Slack is chat, Twitter is chat, email is chat with non-deterministic delays. It replaces slow, synchronous communication with asynchronous, parallelizable communication: just ask any teenager. Moreover, its seemingly simple interface is the HTTP *and* browser for mobile: a protocol and an interface. For instance,

<sup>1</sup><http://money.cnn.com/2016/05/24/technology/snapchat-valuation-20-billion/>

<sup>2</sup><https://www.techinasia.com/talk/wechat-valued-at-83-6-billion-half-of-tencents-market-cap>

<sup>3</sup><http://qz.com/235215/line-ipo-na-ver-will-finally-put-a-real-value-on-messaging-apps/>

<sup>4</sup><http://blog.memsql.com/incumbents-and-contenders-in-the-33b-database-market/>

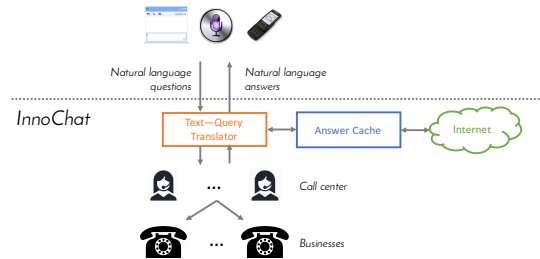


Figure 2: InnoChat System Architecture.

services<sup>5</sup>, assistants<sup>6</sup>, and customer communication<sup>7</sup> can all be accessed through a chat interface. In other words, chat-based accessibility is what having a website was in 1999: innovation.

To this end, chat companies are furiously building APIs and chat-bot frameworks<sup>8</sup> so that forward thinking services can develop services against these APIs. However, most businesses don't have a website, and those with a chat-interface is approximately zero.

Enter the database community. There were 42M registered businesses in the world in 2015, and the majority are connected—not to the internet, but to a telephone. Forget struggling to develop chat-bots for individual businesses: our community can bring a chat interface to every phone-connected business! With only a healthy mix of human-powered computing, information extraction, caching, and data integration.

This abstract presents InnoChat (Figure 2). Natural language questions aimed at phone-connected businesses (e.g., “do any Berkeley hotels have pillows for my British Shorthair cat?”) are submitted through chat boxes, voice, telephones, and other mediums. The Text-Query translator extracts a structure query and attempts to answer it using an existing answer cache as well as the internet. If the cache lookup fails, operators are standing by to directly call the requisite businesses with the question. The phone calls are logged and the answers extracted, inserted into the cache, and returned to the user in chat, voice, or multimedia form. Although deceptively similar to classic crowd-sourced Q&A, InnoChat requires innovations in 1) translating NLP to structured queries, 2) batching and managing call center operators, 3) developing new phone-based, rather than mechanical turk-based question interfaces, and 4) managing the large amounts of success that such a system will bring to the research community.

References available after publication or over chat.

<sup>5</sup><http://a16z.com/2015/08/06/wechat-china-mobile-first/>

<sup>6</sup><https://getmagicnow.com/>

<sup>7</sup><https://www.intercom.io/>

<sup>8</sup><https://dev.botframework.com/>

<sup>9</sup><https://aws.amazon.com/blogs/compute/create-and-deploy-a-chat-bot-to-aws-lambda-in-five-minutes/>