Cognitive Computing and Big Data Analytics
By Judith Hurwitz, Marcia Kaufman and Adrian Bowles

CHAPTER
11

Building a Cognitive Healthcare Application

The healthcare industry is a large and complex ecosystem that encompasses many different types of organizations that support patient wellness and care. The ecosystem is broad, with a number of well-defined roles including:

- Healthcare providers
- Healthcare payers
- Medical device manufacturers
- Pharmaceutical firms
- Independent research labs
- Health information providers
- Government regulatory agencies

Although there have been enormous technological advances that have enabled organizations to improve health outcomes for patients, the need for continued technical innovation is at a tipping point. Each segment of this ecosystem has typically managed healthcare information in a siloed way making it difficult to share patient and medical research data across the various stakeholders. The volume and variety of healthcare data that needs to be managed, analyzed, shared, and secured is growing at a fast pace. Even when participants are motivated to share information for mutual benefit, the required data is often inconsistent and disconnected, which can slow down progress in medical research and lead to clinical errors that put people's lives at risk. Depending on the methodology used to measure medical mistakes, preventable harm leading to death is either the third leading cause of death in the United States behind heart disease and cancer or the sixth behind accidents and ahead of Alzheimer's disease.

This chapter looks at several healthcare organizations where experts are in the early stages of building cognitive applications that help them to solve well-known healthcare problems in new ways, and begin to solve what was previously intractable. These stakeholders in the healthcare ecosystem are beginning to use cognitive systems to help them find patterns and outliers in data that can help to fast track new treatments, improve efficiencies, and treat patients more effectively.

Foundations of Cognitive Computing for Healthcare

The healthcare ecosystem creates and manages a huge volume of data such as digital images from CT scans and MRIs, reports from medical devices, patient medical records, clinical trial results, and billing records. This data exists in many different formats ranging from manual paper records and spreadsheets to unstructured, structured, and streaming data managed in a variety of systems. Some of these systems are well integrated, but most are not. As a result, the vast amount of data generated and analyzed by the healthcare industry presents significant challenges. However, as organizations find new ways to manage and share this data, they are finding that there are amazing opportunities for improving health outcomes. For example, healthcare providers have implemented electronic medical record (EMR) systems to maintain integrated, consistent, and accurate patient records that can be shared by a medical team. Although the EMR is still a work in progress for many organizations, there are great benefits to having a complete, accurate, and up-to-date set of problems and treatment for each patient. Treatment decisions can be made more confidently and with greater speed if the medical information is available in a consistent and accurate form.

One of the persistent challenges for healthcare organizations is the need to find the patterns and outliers in both structured and unstructured data that can help them improve patient care. As shown in Figure 11-1, data management in the healthcare ecosystem is moving away from document-centric silos to well-integrated knowledge bases that include both structured and unstructured data.

The management of healthcare data will begin to follow a more standards-based approach to facilitate sharing of data where appropriate. Medical devices and sensors have the capability to generate valuable data about a patient's condition, but this data is not always captured effectively. There are great opportunities to improve screening of patients and anticipate changes in their medical condition by using predictive analytical models on the data streams. Cognitive systems can capture and integrate this new generation of sensor-based data with the entire recorded history of medical research and clinical outcomes captured in natural language text to form a corpus. The system learns from experiences with this corpus, enabling significant improvements in outcomes.

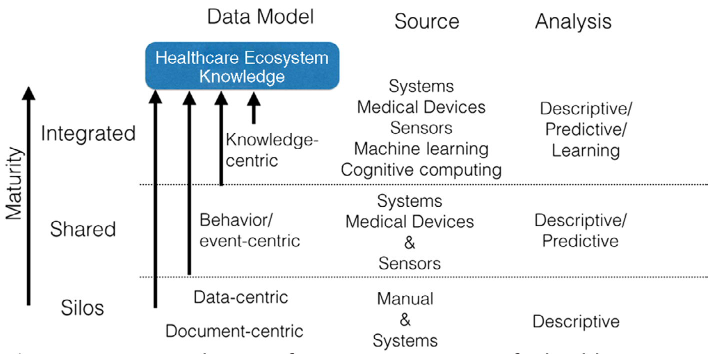

Figure 11-1: Foundations of cognitive computing for healthcare

For example, doctors in the neonatal department of a Toronto hospital developed analytical models that provide 24 hours of advance warning on which babies might develop a life-threatening infection. The infection, late-onset neonatal sepsis, is a blood infection that can occur in subsets of newborn babies. Prior to the analytics research done by Dr. Carolyn McGregor, Canada Research Chair in Health Informatics based at the University of Ontario Institute of Technology, the neonatal intensive care unit relied on monitors that collected data on infants' vital signs but stored only 24 hours of data at a time. By capturing the data from the monitors as an ongoing stream of data, the informatics team developed algorithms to analyze the data over time. The algorithm looks for patterns that occur before the infection becomes clinically apparent. With the new system doctors get a digital reading on respiratory rates, heart rates, blood pressure, and blood oxygen saturation, and can monitor infants' vital signs in real time and detect changes in their conditions.

Constituents in the Healthcare Ecosystem

The healthcare ecosystem has evolved to include a variety of organizations, each of which contributes to the development, financing or delivery of wellness or treatment information, processes, or products. As shown in Figure 11-2, healthcare providers, payers, pharmaceutical companies, independent research groups, data service providers, and medical manufacturers all have access to different sources of relevant healthcare data. Government agencies and even the patients have a role in managing who sees which data. Some of this data is shared, but much of it is controlled by regulations and security requirements. The relationships between the constituents in terms of data sharing are complex and in a state of flux. To move toward a more integrated approach to healthcare ecosystem knowledge that includes more predictive analysis and machine learning, there needs to be continued improvement in the consistency of data shared across the ecosystem.

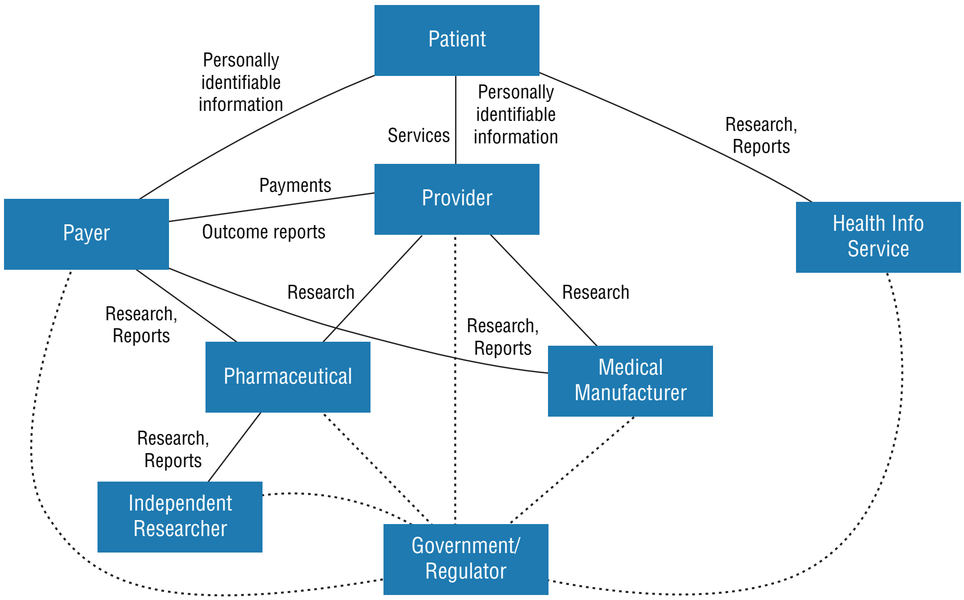

Figure 11-2: Healthcare ecosystems data sources

The data managed and leveraged by different constituents in the healthcare ecosystem includes:

- **Patients**—From family history and habits to test results, individuals participating in the healthcare ecosystem produce personally identifiable information, which may be aggregated anonymously, where permitted, to guide care for those with similar attributes.
- **Providers**—Data covers a broad range of unstructured and structured sources. Some examples include patient medical records (EMR, doctors' office notes, and lab data), data from sensors and medical devices, intake records from the hospital, medical text books, journal articles, clinical research studies, regulatory reports, billing data, and operational expense data.
- **Pharmaceutical companies**—Data to support pharmaceutical research, clinical trials, drug effectiveness, competitive data, and drug prescriptions by medical providers.

- **Payers**—Data includes billing data and utilization review data.
- **Government agencies**—Regulatory data.
- **Data service providers**—Prescription drug usage and effectiveness data, healthcare terminology taxonomies, and software solutions to analyze healthcare data.

Learning from Patterns in Healthcare Data

The benefit of cognitive computing is that healthcare professionals will more easily get the insights they need from all types of data and content to act with confidence and optimize their decision making. The risks of not finding the right relationships and patterns in the data are high in the healthcare industry. In this industry, if important pieces of information are overlooked or misunderstood, patients can suffer long-term harm or even death. By combining technologies such as machine learning, artificial intelligence, and natural language processing, cognitive computing can help healthcare professionals to learn from patterns and relationships discovered in data. The collaboration between human and machine that is inherent in a cognitive system supports a best practices approach that enables healthcare organizations to gain more value from data and solve complex problems.

Gaining more value from data is a multifaceted process that requires both technology and human knowledge. Getting the data right is paramount. The relevant data needs to be accurate, trusted, consistent, and available for access expeditiously. However, having accurate data is only the baseline for improving health outcomes for patients. Physicians need the skill and experience to make sense out of what is often a complex set of symptoms and diagnostic tests. They need to internalize best practices that enable them to ask the right questions and listen for answers from the patient. The solution to a patient's problem is not always obvious in the medical lab results and images. Best practices that focus on connecting all the disparate data points can help physicians, researchers, and others in the healthcare ecosystem to find the right solution.

Learning from patterns in data helps healthcare organizations to solve some of their most challenging problems. For example, The University of Iowa Hospitals and Clinics has identified patterns in a population of surgical patients that help to improve both quality and performance in surgery. The hospital has modeled data from hospital readmission, surgical site infections, and other hospital-acquired infections. The model enables physicians to predict which patients are most at risk for acquiring a surgical site infection while they are still in the operating room and corrective actions can be taken.

Other hospitals use predictive models to reduce costly and dangerous hospital readmission rates as well. The patterns identified from thousands of hospital records are used to build a model that can analyze a patient's medical record to identify risk factors for problems that may occur after discharge from the hospital. Predictive analytics models look at a number of different factors to determine which ones have the greatest impact on hospital readmission rates. As shown in Table 11-1, these factors may be specific to the patient or the physician.

Table 11-1: Attributes to Consider for a Predictive Model on Hospital Readmissions

Patient Attributes	Smoker, drug abuse, alcohol abuse, lives alone, dietary noncompliance
Socio-economic Attributes	Educational status, financial status
Physician Factors	Incorrect medicines given, overlooked important information about patient

Understanding the risk factors can help hospitals to improve processes within the hospital and take corrective measures to decrease hospital readmission rates. The predictive model can help on a case-by case basis by indicating which patients may require more intensive follow up after they are discharged.

Building on a Foundation of Big Data Analytics

Although there is a great deal of interest and some exciting case examples of cognitive systems in healthcare, these implementations are at an early stage. However, healthcare organizations are not starting from scratch for cognitive computing and big data analytics. There are some high-profile examples of analyzing data and incorporating machine learning in medical environments. The next generation of these healthcare platforms are building on a strong foundation of big data analytics. As healthcare informatics capabilities mature to incorporate cognitive systems, the overall goals for the healthcare organization remain the same. There is a common focus on providing optimal high-quality care to patients and to continually improve healthcare options and outcomes in a cost-effective manner.

Much of the effort in healthcare IT has been focused on developing more integrated systems so that medical information can be safely stored and accessed as needed for research and patient care. For example, healthcare providers have implemented electronic medical records (EMR) to help provide a unified record of medical data for each patient. Much of the patient-related data is unstructured, and large volumes of this data come from digital images, lab tests, pathology reports, and physician reports. As described in the previous section, healthcare organizations are rapidly finding new ways to gain value from this data. In addition to using EMR and other patient-specific data to make decisions for one patient, there is great value in leveraging data across large groups of patients to build predictive models that can improve outcomes for large populations of patients. These analytics efforts need to ensure that requirements for security and privacy are met by removing any personal identifying information from the data.

One healthcare field where big data analytics is rapidly increasing the speed at which new research can be completed is the biopharmaceutical field. Revolutionary advances in DNA sequencing technology makes it possible to collect huge volumes of genomic information for analysis. To keep up the pace of the research, technology is used for sequencing data storage, processing, and downstream analytics. There are huge demands for new computational approaches to store and analyze genomic data. Advanced algorithms, methods, and tools enable scientists to effectively understand the data produced by genomic analyses, and to help answer important biological questions. Advanced modeling efforts are replacing many of the more manual efforts used in the past to analyze genomic data.

Cognitive Applications across the Healthcare Ecosystem

Many healthcare experts are building on what they have already achieved in big data and analytics initiatives to incorporate machine learning and cognitive computing. The goal is to continue to optimize results in healthcare research and clinical diagnosis and treatment. Gains in speed, innovation, and the quality of outcomes are dependent on how humans interact with the available technology and data. In addition, there is an exceedingly strong requirement within healthcare organizations for those humans with the most experience to put best practices in action and to share those best practices with the next generation of healthcare professionals. This transfer of knowledge takes place continuously through training programs for medical students and residents as well as assistant and mentoring programs in research labs. Introducing a cognitive system to support healthcare professionals as they learn can help drive this process of knowledge transfer forward. Right now the use of cognitive computing is at an early stage; however, the expectation is that over the next decade it will become well integrated into many healthcare processes.

Two Different Approaches to Emerging Cognitive Healthcare Applications

The implementations of cognitive healthcare applications are proceeding along two different paths: customer or user engagement applications and discovery applications. Customer engagement applications are designed to help find personalized answers to questions. For example, several emerging companies have developed cognitive applications that provide consumers with answers to questions about managing their own health and wellness. Other cognitive systems provide support for healthcare payer customer service agents. With a corpus that contains more relevant information than people could possibly consume and retain, these systems answer relevant questions and provide new insight about their health. Discovery applications are used in situations such as drug discovery or to discover the optimal treatment for a patient. In both types of cases, the healthcare organization needs to begin by defining the end user of the system, the types of questions that will be asked, and the content that is required to build the knowledge base for the system. The cognitive system is used to understand relationships and discover patterns in data that may lead to improved healthcare outcomes.

You need to understand the types of users who access your cognitive healthcare application. What is the medical background and expertise of your users? For example, will the users be medical students, or medical clinicians with many years of experience? Or will your users be health and wellness consumers? Expectations for user/system interaction will have an impact on the development of the corpus, the design of the user interface, and how the system is trained. The user type also has an impact on the confidence levels required and level of accuracy the system needs to achieve. As user requirements and expectations change over time, these changes must be incorporated into the ongoing development of the cognitive system. The learning process for a cognitive system is continuous, and as a result the system gets smarter and delivers greater value to end users the more it is used.

The Role of Healthcare Ontologies in a Cognitive Application

Healthcare taxonomies and ontologies—a coding system or semantic network of medical terms and the relationships among these terms—are important to the development of a corpus for cognitive healthcare applications. These ontologies are used to map the relationships between terms with similar meanings. There are many ontologies that are already widely used in healthcare to organize terminology related to medical conditions, medical treatments, diagnostic tests, ingredients and dosing for clinical drugs, and drug complications. One example of a medical ontology is the International Classification of Diseases (ICD). The ICD-10 is the current version as endorsed by the World Health Organization. However, it has not yet become the standard in all countries. ICD-10 will become the standard in the United States some time after October 1, 2015. The ICD includes codes for diseases, disease symptoms, and medical findings about diseases. The ICD is only one of many different taxonomies and ontologies in use across the ecosystem. To build an efficient corpus for a healthcare application, you need to find a common language to ensure that data from different sources can be integrated and shared. Without a taxonomy of terms, the cognitive system cannot learn as quickly, and the accuracy of results will be insufficient. Your systems will miss a lot of terms that have the same meaning.

Healthline Corporation has developed one of the largest semantic taxonomies for the healthcare ecosystem. It maps the relationships between consumer and clinical applications, which can help to support new consumer-focused cognitive health applications. Algorithms can reference the taxonomy to improve the semantic understanding of a query to the cognitive system. In addition, cognitive health applications can make more accurate associations between medical concepts by referencing a comprehensive and accurate ontology or taxonomy.

Starting with a Cognitive Application for Healthcare

Early stage examples of cognitive applications in healthcare are built on top of the cognitive engine or platform. To develop an application you need to begin by defining your target end user and then train the cognitive system to meet the needs of your user base. What is the general subject area for your cognitive application? What do you know about your users' level of knowledge in this area, and what are their expectations or requirements from the cognitive application?

A cognitive system needs to start with a base level of information from which it can begin to find the linkages and patterns that can help it to learn. Although the learning process begins with questions, a trained system can do much more than provide answers to a set of questions. The cognitive system can make associations between questions, answers, and content to help the user understand the subject matter at a deeper level. The basic steps required to build a cognitive application in healthcare follow.

Define the Questions Users will Ask

You want to begin by assembling the types of questions that will be asked by a representative group of users. After this step is completed, you can assemble the knowledge base required to answer the questions and train the system effectively. Although you may be tempted to begin by reviewing data sources so that you can build your knowledge base or corpus for your system, best practices indicate that you need to take a step back and define your overall application strategy. The risk to beginning with your corpus is that you are likely to target your questions to the sources you have assembled. If you begin with the corpus, you may find you cannot meet the needs of your end users when you move to an operational state.

These initial questions need to represent the various types of questions users will ask. What do users want to ask and how will they ask questions? Are you building a consumer-focused application that will be used by a general population of users, or are you developing a system that will be used by technical experts? Getting the questions right is critical to the future performance of the application. You need to seed the cognitive system with a sufficient number of question and answer pairs to start the machine learning process. Typically, 1000–2000 question/answer pairs seem to be the right number to get the process started. Although the questions need to be in the voice of the end user of the system, the answers need to be determined by subject matter experts.

Ingest Content to Create the Corpus

The corpus provides the base of knowledge used by the cognitive application to answer questions and provide responses to queries. All the documents the cognitive application needs to access will be included in the corpus. The question/answer pairs you have created help to drive the process of collecting the content. By beginning with the questions, you have a better idea of the content that will be required to build the corpus. What content do you need to answer the questions accurately? You need to identify the resources you have and which resources you may need to acquire to provide the right knowledge base. Examples of content include medical texts, background information on health subjects such as pharmaceutical research, clinical studies, and nutrition, medical journal articles, patient records, and ontologies and taxonomies.

The content you select needs to be validated to ensure that it is readable and comprehensible. Adding meta tags to your content can help with creating associations between documents. For example, you can use tagging to identify that an article pertains to a specific medical condition such as diabetes. In addition, content should have sections and headings to provide cues to the cognitive system. You may need to optimize the format of some of the source data to ensure that it can be properly identified and searched. For example, structured data sources such as a comprehensive nutrition table may need to be transformed into unstructured content prior to ingestion into the corpus. Simple tables can be read by the cognitive system, but more complex and nested tables should be transformed to unstructured text for clarity. The source transformation process is required to ensure that the corpus functions properly.

You need to understand the life cycle of documents you ingest to plan for appropriately scheduled updates. In addition, you may need to establish a process that will ensure you are notified of new and updated content. The corpus needs to be updated continuously throughout the life of the application to make sure it continues to be viable.

Training the Cognitive System

How does the training process begin? The cognitive system learns through analysis and training (refer to Chapter 1, "The Foundation of Cognitive Computing," for discussion on different types of machine learning and Chapter 3, "Natural Language Processing in Support of a Cognitive System," for details on natural language processing). Just think of how you might approach learning a new subject. Initially you may have a long list of questions. You do some reading and then your questions change in content and scope as you learn more about the subject. The more you read and understand, the fewer questions you have. A cognitive system is similar in that the more question/answer pairs that are analyzed, the more the system learns and understands.

Analyzing the question/answer pairs is a key part of the overall training process. Although it is important for representative users to generate the questions, experts need to generate the answers and finalize the question/answer pairs. The questions need to be consistent with the level of knowledge of the end user. However, the experts need to ensure that the answers are accurate and in line with the content in the corpus. As shown in Table 11-2, you are likely to have some overlapping questions or clusters of questions. These questions may ask about a similar topic using slightly different terms or from a different perspective. Or the questions may be basically the same except one version of the question abbreviates certain terms. The cognitive system learns from these clusters of questions.

Table 11-2: Questions Used to Train a Cognitive Application on Health and Wellness

Question 1	What is the difference between whole milk and skim milk?
Question 2	Is low fat milk different from whole milk?
Question 3	Is skim milk better than whole milk?

Question Enrichment and Adding to the Corpus

The training process for your cognitive application is used to ensure your application works as intended when it becomes operational. Initially, it needs to be repeated multiple times using training data, test data, and blind test data. As each of these tests is completed, you can add content to the corpus to cover areas in which there is inadequate information.

Plan to continually return to the training process after your application goes live so that you establish an ongoing process of updating question/answer pairs and adding to the corpus. Expansion algorithms are used to determine which additional information would do the best job of filling in gaps and adding nuance to the information sources in the corpus.

Using Cognitive Applications to Improve Health and Wellness

The patient (or healthcare consumer) is central to the healthcare ecosystem (refer to Figure 11-2). This complex ecosystem generates an enormous amount of data that describes the health and well-being of every individual in the system. Many organizations that manage a population of healthcare consumers have implemented various programs to help improve the group's overall health. The challenge is that these programs do not always provide the personalized responses and incentives that their members need to change behavior and optimize health outcomes. The payback of helping individuals to lose weight, increase exercise, eat a well-balanced diet, stop smoking, and make healthy choices overall is huge. Healthcare payers, governments, and organizations all benefit if communities as a whole are healthier and individuals do a better job of managing previously diagnosed conditions. The following list (developed by the Office of the Surgeon General (U.S.); Office of Disease Prevention and Health Promotion (U.S.); Centers for Disease Control and Prevention (U.S.); National Institutes of Health (U.S.); and the Rockville (MD, U.S.) Office of the Surgeon General (U.S.); 2001) shows the many different medical conditions and diseases increased weight is associated with. Even when faced with these facts, it is incredibly hard for many people to make the positive change they need. These conditions and diseases include:

- Premature death
- Type 2 diabetes
- Heart disease
- Stroke
- Hypertension
- Gallbladder disease
- Osteoarthritis
- Sleep apnea
- Asthma and other breathing problems
- Certain types of cancer
- High cholesterol

Finding ways to improve the connections and communication of individuals and the healthcare ecosystem is a priority for a number of emerging companies. Several of these companies are highlighted in the following sections.

Welltok

Welltok, based in Denver, CO, provides personalized information and social support to help individuals optimize their health with its CaféWell Health Optimization Platform. Welltok works with population health managers, such as health payers, to help decrease healthcare costs by providing a platform that gives people the support, education, and incentives (e.g., gift cards, premium reductions) they need to change their behavior and improve their health.

Overview of Welltok's Solution

Welltok's CaféWell Concierge is a platform designed to help individuals optimize their health by connecting them with the right resources and programs. It organizes the growing spectrum of health and condition management programs and resources, such as tracking devices, apps, and communities, and creates personalized, adaptive plans for each consumer.

Welltok partnered with IBM Watson to create the CaféWell Concierge app, which leverages cognitive technologies to dialogue with consumers and provide personalized guidance to optimize their health. Vast amounts of internal and external data sources are used to build corpora that form the knowledge base of the system. CaféWell Concierge uses natural language processing, machine learning, and analytics to provide personalized and accurate recommendations and answers to questions asked by individual health consumers.

As a mobile application, health consumers can engage with the CaféWell Concierge at a time and place that is convenient for them. Each individual receives an Intelligent Health Itinerary based on their health benefits, health status, preferences, interests, demographics, and other factors. The itinerary is a personalized action plan with resources, activities, health content, and condition management programs. For example, consumers with controllable health conditions like diabetes or asthma will receive an Intelligent Health Itinerary with educational information and guidelines tailored to them to help them make healthy choices on a daily basis.

Welltok's partners include health payers that make the app available for free to their members. The health payers typically offer incentives or rewards such as entry in a drawing for a gift card for completing a coaching session, or a reduction in health costs for improving your body mass index (BMI). CaféWell uses advanced analytics algorithms to align actions and behaviors with the right incentives and rewards to motivate consumers to get involved in their health. It also learns over time what individuals respond to and what type of incentives or value to offer for targeted behaviors.

Using natural language processing, consumers can dialogue with the application and ask questions related to health and wellness. Welltok followed the steps in the previous section to build a cognitive application that can handle mass personalization, process large volumes of information, and answer open-ended questions in seconds. The architecture and data flow for the question-answer training process for CaféWell is illustrated in Figure 11-3.

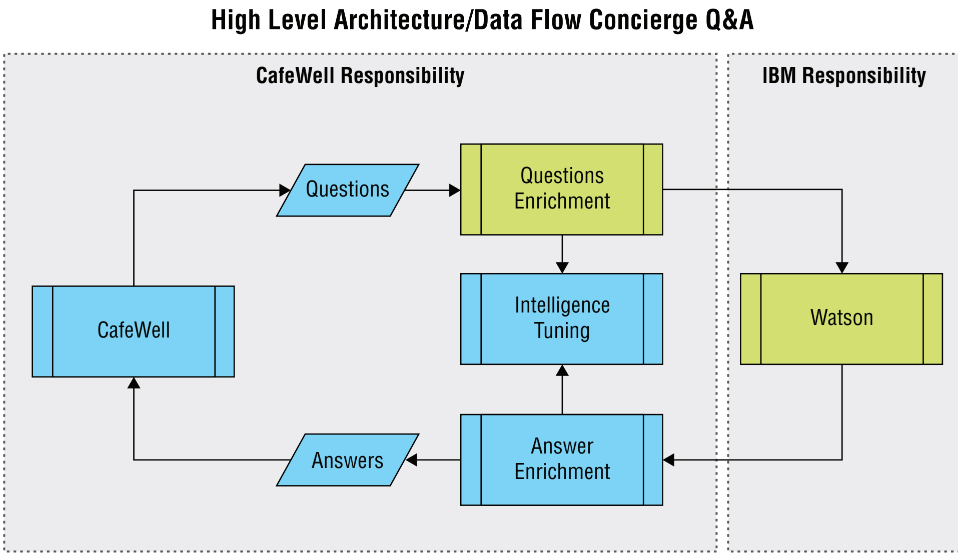

Figure 11-3: Welltok training architecture

To develop its question/answer pairs, Welltok collected input from consumers to create questions that would reflect their interests and used subject matter experts to answer the questions logically and accurately. Table 11-3 shows a sample of the thousands of question/answer pairs that Welltok created to begin the training process for CaféWell Concierge. After determining an initial set of question/answer pairs, Welltok developed the corpora (and ontologies) for the application to provide Watson with access to the information sources it needs. Welltok collected unstructured information from third-party healthcare sources to get all the information required for the corpora.

Welltok worked closely with IBM to train Watson for CaféWell Concierge. The iterative process of ingesting data to build the corpus, enrich content, and improve the intelligence of the cognitive system is illustrated in Figure 11-4. By leveraging Watson's cognitive capabilities, CaféWell can understand context and learn about a user's health concerns, goals, and preferences. Watson's machine learning capabilities enables CaféWell to continuously improve the quality of its responses and recommendations. Watson has dozens of different corpora covering many different aspects of health and wellness, including health insurance benefits, nutrition, and fitness. These corpora, in addition to information about the individual, are used to support the advanced analytics algorithms that deliver the personalized recommendations and responses. The application goes beyond providing search results. It builds a relationship with the users—getting to know them, providing personalized recommendations and guiding them to optimal health.

Table 11-3: Sample of Welltok Question/Answer Pairs

What are some lifestyle changes that I should make if I have high blood pressure?	Lifestyle changes are just as important as taking medications. Reducing your weight by just 10 pounds may be enough to lower your blood pressure. Losing weight can help to enhance the effects of high blood pressure medication and may also reduce other risk factors, such as diabetes and high bad cholesterol.
How do you determine the calories burned by your body?	BMR is often calculated using the Harris–Benedict equation. This equation calculates basal metabolic rate based off of 3 variables: weight, height, and age. Using this approach, total energy expenditure can be calculated by multiplying BMR by an activity factor. **Equation For Men:** BMR = 88.362 + (13.397 × weight in kg) + (4.799 × height in cm) – (5.677 × age in years)
Do my nutritional needs vary throughout life?	Nutritional needs vary throughout life. From infancy through adulthood, good nutrition is essential to growth and development, and to maintaining health in the later years.
Why should I read the food label on packaged foods?	Most packaged foods have a label listing nutrition facts and an ingredient list. In the U.S., the Food and Drug Administration (FDA) oversees the requirements and design of the Nutrition Facts label. The purpose of the label is to help consumers make quick, informed food choices that contribute to a healthy diet. Especially on a low-sodium diet, you need to look at the food label to limit sodium intake.
I have a grain allergy, what food should I avoid? What kinds of food are considered grains?	Any food made from wheat, rice, oats, cornmeal, barley, or another cereal grain is a grain product. Bread, pasta, oatmeal, breakfast cereals, tortillas, and grits are examples of grain products. There are whole grains, containing the grain kernel, and refined grains, which have been milled to remove bran and germ. There are many benefits to a diet rich in grains.

Using Watson's machine learning capabilities, CaféWell Concierge improves the quality of responses users receive with each interaction. And with its spatial and temporal capabilities, the application factors in time and location to provide highly relevant information. For example, it can recommend where and what to eat for lunch based on your location and your specific diet and nutritional requirements.

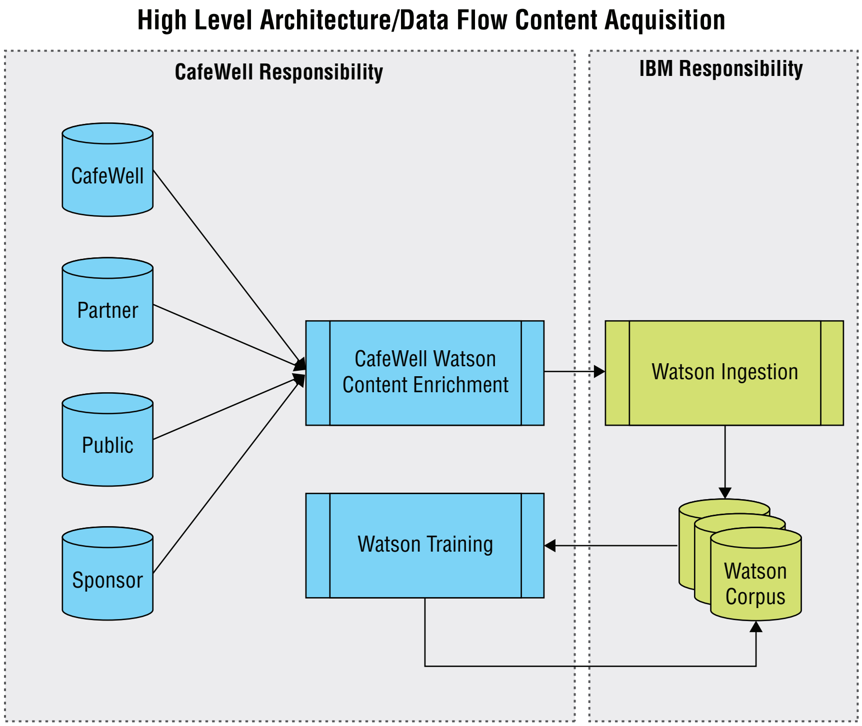

Figure 11-4: Welltok high-level architecture and data flow: data flow content acquisition

CaféWell Concierge in Action

CaféWell Concierge is intended to help individuals understand their health status and receive personalized guidance to help them achieve wanted health and wellness goals, and get rewarded for doing so. The following example shows how an individual with a new medical diagnosis could benefit from interacting with CafeWell Concierge.

Assume you just received a new diagnosis of pre-diabetes from your doctor. You saw your internist in his office last week and after examining you he took some tests. Today, you received a follow-up phone call with your diagnosis and a recommendation to change your diet, lose 20 pounds, and increase your exercise. However, you travel a lot for work leading you to eat a lot of meals at restaurants, and you never have enough time to get to the gym. What do you do next?

Without cognitive computing you might search the Internet for Type 2 diabetes and find information that leaves you confused and scared. Although there are many applications that allow you to search for nutritional information and monitor your weight and exercise activity, they only provide general information about pre-diabetes. A cognitive application can provide you with deeper insight and more personalized high-quality support. CaféWell Concierge creates an Intelligent Health Itinerary for you including programs and resources such as a video coaching session on nutrition, food choices at local restaurants, a fitness tracking device with step goals to help reduce your BMI, and a social community for additional support.

Based on Watson's cognitive capabilities, CaféWell Concierge can integrate and analyze across multiple sources of information so that you can receive tailored and continuous support as you might from a personal concierge.

GenieMD

GenieMD is also focused on providing a cognitive health application for consumers. The company's mission is to help clients develop more meaningful conversations with their healthcare providers. The overall goal is to help health consumers take a more active role in managing their own health and the health of their loved ones. Users can ask questions in natural language and receive personalized responses and recommendations. These users can access GenieMD through a mobile application. The expectation is that patients will achieve improved health outcomes and that healthcare costs will be lowered. GenieMD aggregates medical information from a range of disconnected sources and makes that information actionable. GenieMD, which is powered by IBM's Watson, is following a development process that is similar to Welltok.

Consumer Health Data Platforms

Google, Apple, and Samsung are all developing consumer focused heath data platforms. These platforms are in the early stage, and the type and variety of data being collected is narrower than the applications discussed in the preceding sections. Google has a set of Google Fit APIs that it provides to developers to help them manage and combine different types of health data. At this point, the health data collected typically comes from wearable devices such as the FitBit, Nike Fuel Band, and other medical sensors that can detect biometric data. This data includes heart rate, steps taken, and blood sugar level. Nike FuelBand can publish user heath data that it collects to the Google Fit platform.

Using a Cognitive Application to Enhance the Electronic Medical Record

The electronic medical record (EMR) is a digital record of the medical and clinical data for each patient followed by a provider (independent physician or large medical centers with all physicians in one group). Typically, the EMR is designed to store and retrieve data about a patient that can be used for diagnosis and treatment. It has some basic reporting capabilities such as flagging a lab test as low or high based on predetermined criteria. The EMR can be thought of as having three main functions: Think, Document, and Act. Today, the EMR documents information about a patient and supports the physician's ability to take actions on a patient's behalf. But, the EMR does not help with the "thinking" aspect of determining how to best deliver care to a patient. By incorporating machine learning, analytics, and cognitive capabilities into the EMR, physicians could be guided in understanding how a diagnosis was arrived at and the issues surrounding a treatment plan. Overall, healthcare organizations would like to gain more value from the EMR including finding ways to leverage the information in the EMR to improve coordination between different providers and providing more individualized high-quality care for patients.

Epic Systems, a healthcare software company providing EMR software, holds the medical records of approximately 50 percent of the patients in the United States. The company partnered with IBM to add a content analysis capability to the EMR. This enables physicians to use text-based information on a patient as part of the electronic medical record. IBM's natural language processing software, IBM Content Analytics, enables physicians to extract insights from the unstructured text in real time. The EMR can be used with a cognitive system that enables physicians to get answers to complex questions about patient diagnosis and treatment. The information stored in the EMR could be either incorporated into the corpus of the cognitive system or used as part of an analytics engine that is integrated with the cognitive system. Epic's approach provides for analysis of the physician's text-based notes on a patient and transforms these notes into a format that can be incorporated into the patient record. By automatically applying industry-standard diagnosis and treatment codes, significant improvements in accuracy and efficiency can be achieved.

Hitachi is working on a number of consulting projects with healthcare organizations related to enhancing the business value achieved from the EMR. In one project, Hitachi is working with a hospital and EMR vendor to help determine if the treatment plan selected for the patient is the best and most cost effective. Hitachi provides a clinical repository with an analytics engine and a database extractor tool. The focus is on gaining value from the unstructured content.

The Cleveland Clinic is working with IBM's Watson on a cognitive healthcare system focused on rethinking the capabilities of the EMR. How can the EMR be more accurate and used to help physicians learn about the thought process behind clinical decisions? Dr. Martin Harris, Cleveland Clinic, explained how important it is to create one unified and accurate problem list for all patients. One patient may see four specialists for different medical conditions, but for the benefit of the patient, there must be one problem list that includes all of that patient's medical information. Any omission from the EMR could lead to incomplete knowledge of the patient's conditions, and the impact to the patient can be dangerous.

Although there is some risk that the EMR may be missing an important piece of information, there is typically a lot of information to review in each patient's record. Nothing is deleted in the EMR, and it can be hard to find the information you are looking for. If a patient has a complicated medical condition, the EMR can easily reach 200 pages or more. Given the volume of information in a patient's EMR, some physicians find it more cumbersome to use than the old paper records.

The Cleveland Clinic is building a comprehensive knowledge base using IBM's Watson that can be used to test for omissions and improve the accuracy of its EMR. The Cleveland Clinic ingested information from the EMR into the corpus of the cognitive system along with unstructured data including physician and hospital admission notes. When the unstructured data is compared to the problem list from the EMR, all too often omissions are identified. Using the cognitive system to ask questions would enable the hospital to make sure they are retrieving all information about a patient when needed for analysis. The goal of this project is to develop an EMR assistant that would provide a visual summary of a patient's condition. Users could type in keywords and receive visualizations that would help research a patient's medical history and improve decision making.

Using a Cognitive Application to Improve Clinical Teaching

The most senior and experienced physicians on the staff of a medical center are responsible for transferring knowledge about clinical diagnosis and treatment to medical students and residents. In addition, senior clinicians and researchers at large teaching medical centers want to share knowledge with staff at smaller community hospitals. Research in many areas such as cancer are advancing so quickly that experts at some of the largest medical teaching centers say that it can take years before the information on the newest treatments is translated to changes in treatment offered at community hospitals. In medicine, one is always a student. Each subspecialty has major conferences across the world where papers on the latest research are presented and medical knowledge is shared. In addition, physicians read journal articles to keep current with new research. One service used by many physicians is UpToDate, a clinical decision support resource that provides edited summaries of recent medical information in addition to evidence-based recommendations for treatment. Even with all these resources, it is extremely challenging to keep up with all the new discoveries in drugs and treatment options.

The training of the next generation of physicians is of the utmost importance to members of a senior medical team. Physician leaders at several top medical institutions are developing cognitive systems that may add new dimensions to the complex task of transferring knowledge about medical best practices and diagnostic skills. The expectation is that these new cognitive systems would be in addition to the traditional methods of personalized instruction followed in teaching medical centers. Physicians who train side by side with senior experts in the field learn lessons that they carry with them throughout their careers. A senior neurologist at a Boston teaching hospital described his role as one of "modeling the behavior that students need to follow in treating patients." A large team of medical students and residents accompany him as he makes rounds in the hospital. Students need to be exposed to a variety of diseases in each subspecialty and learn how to identify the differential diagnosis based on symptoms. However, his teaching goes much deeper than understanding the symptoms and treatments of diseases. He wants the students and residents to learn what questions to ask and how to ask them to get the information they need to deliver optimal care.

The Cleveland Clinic is working with IBM to develop a cognitive system called Watson Paths that will provide additional knowledge to students to support what they learn on their subspecialty rotations. Typically, students rotate through a series of subspecialty rotations, spending one month or more in each rotation. Students' clinical experience varies depending on the cases in the hospital during the time they are in the subspecialty unit in the hospital. Cognitive systems that have been well trained in how to treat a broad spectrum of diseases can change the way medical students are taught.

If a medical student is exposed to the cognitive system before the rotation, then the whole training process can become more powerful and can go deeper. If the student has a better understanding of the process of making a diagnosis on some of the most common conditions, the attending physician can focus more on the less obvious diagnosis. The focus needs to be on having as much information as possible to make a correct diagnosis. Considering that there are approximately 13,000 diagnosis codes in the ICD-9 and more than 68,000 diagnosis codes in the ICD-10, students have a lot to learn. A good internist may know approximately 600 medical diagnoses, whereas a subspecialist may have deep knowledge in 60 diagnoses. Fortunately, a cognitive system can ingest information on a huge scale. Cognitive systems (after training) can produce scenarios for the top 600 diagnoses and provide guidance to medical students to help them learn by showing the step-by-step approach in making a diagnosis. As cognitive systems can keep track of the evidence used to support their hypotheses and conclusions, they can justify the resulting confidence level that the diagnosis is right.

Students will be able to interact with Watson Paths to study different approaches to treating patients with a certain set of problems. The students can interact with a system that offers reference graphs and probabilities of outcomes depending on the treatment approach the doctor and patient decide to follow. Watson Paths will focus on evidence-based learning: validating and calibrating the impact of treatment options that are selected. The system will annotate each decision—helping students to learn the impact of their decisions. As a result of its machine learning capabilities, the more people who interact with Watson paths, the better it gets in accuracy and understanding.

Memorial Sloan Kettering (MSK) is also working with IBM to develop a medical cognitive system powered by Watson. MSK is one of the top cancer research and treatment centers in the world, and its physician leaders are concerned about the length of time it takes for new research to reach the thousands of medical and surgical oncologists that are not based at one of the large cancer centers. MSK identifies the sharing of medical knowledge about cancer diagnosis and treatment as an important part of its mission. The medical center has more than 30 physicians working on the initiative to ingest data from a huge patient database and train Watson.

There is often more than one approach that will work to treat a particular cancer patient. MSK is helping to train Watson so that a physician can get help in assessing the potential outcomes of using one approach versus another. The expectation is that the oncology cognitive system will help to increase the speed at which new treatments can be disseminated. Watson will help with suggestions and support the physician who needs to make decisions about the best approach for his patient.

Summary

We stated previously in this chapter that it is early for cognitive healthcare applications. It is not easy to project how quickly these applications will evolve and become more integrated with operations across the healthcare ecosystem. However, the significant partnerships that are rapidly forming between healthcare experts and technology leaders in cognitive computing suggest a rapid increase in the pace of development. There are many reasons for the increasing investment of time and money in the development of cognitive healthcare applications. However, the most powerful driver for developing cognitive healthcare applications stems from the challenge of gaining insight from the large and rapidly growing volumes of structured and unstructured data managed by the healthcare ecosystem. There is an overabundance of data generated by the healthcare ecosystem that is not well integrated and not easily shared.

Many of the initial healthcare cognitive computing efforts are focused on how patients engage with their own data. Welltok's Café Well Concierge and GenieMD are excellent examples of this type of application. These applications focus on how the patient communicates with healthcare providers and gains access to information about their medical conditions in a meaningful way. These are practical applications that can help a health consumer adjust his priorities for diet and exercise to improve his overall health. On the other end of the spectrum, there are some interesting applications focused on what can be learned from medical best practices. Clinicians and researchers make decisions that impact people's lives on a daily basis. All too often these decisions are made without the knowledge that comes from a comprehensive understanding of best practices. The goal of many of these emerging cognitive health applications is to ensure that all physicians have the opportunity to evaluate their clinical diagnosis and treatment options in collaboration with a well-trained cognitive system.