Is Natural Language Processing the key to data-driven health care?
Why the need for data are rising

The need for clinical data has never been greater in health care. Data are needed to improve the quality of care, to meet Meaningful Use requirements, to prepare for the ICD-10 transition, to effectively manage populations, and to enable clinical research – all are data-driven.

And as the shift to capitated payment models draws near and accountable care becomes the new reality, the need for data is accelerating. In a recent survey of CHIME executives, “…more than 90 percent of respondents said analytics will be ‘extremely important’ or ‘very important’ to their organization within the next 1-3 years.”

Unfortunately, there is an inherent lack of ‘usable’ data in health care today. Often, facilities have invested in numerous (disparate) systems to conduct particular tasks, like coding or CDI, under the guise of ‘best of breed’ solutions. A product for this, a product for that… it’s the ‘best for our needs.’ And as a result, healthcare IT is so inundated trying to maintain the mass of hardware, systems and interfaces to support this approach that taking on new projects seems untenable. Then one day, maybe recently, it becomes evident that there is no singular source of data to make decisions from, and while that wasn’t an issue in years past, it is becoming clearer and clearer that it will be soon. Without good data there is an information deficit.

Many believe the EHR will become that singular source for their clinical data, and generate the information needed. Not likely. If your EHR contains free text – and it undoubtedly does – that text is unstructured and is not data. It is text… unusable text. And as stated in the CHIME survey, “other important initiatives ranked by respondents included the need... for a ‘single version of the truth’ (59 percent).”

Finding data, for most, is like finding a “needle in a haystack”, and the EHR is now the haystack! If you can’t find data, you can’t use data. How will you identify ‘the patients’ you need to notice?
The documentation conundrum

If you are in the majority, there is valuable clinical information that is trapped in the text of your paper notes, in your dictated and reports, and in your electronic systems – including your EHR. The data in the EHR comes from the structured entry fields, like Problem Lists, that the clinicians (perhaps reluctantly) key in. Often, though not always, these lists are powered by terminology technologies like those from Intelligent Medical Objects (IMO) – the industry leader. Those lists are recorded as structured data.

The data challenge is not as much the EHR as the free text, of which much is ever likely to convert to the aforementioned data entry techniques. The most notable challenge is the dictated reports created by radiologists, pathologists and surgeons. Few believe these reports will ever be created via structured data entry.

The $64,000 question then is:

*How do you generate quality, normalized structured data from all the various sources without disrupting or modifying physician behavior and use the information to improve care quality and reimbursement?*

Many believe the answer lies in Natural Language Processing (NLP). Do you? Or are you not convinced? This white paper explores the appropriate use of NLP in generating data and using information and addresses the question:

**Is NLP alone sufficient?**

Why NLP alone is insufficient

Natural language processing is garnering more and more attention in the healthcare industry of late. The need for facilities to create and utilize data has created an extraordinary focus on the ability for NLP to generate well-formed clinical data. It seems that the focus surrounding NLP has been centered upon the efficacy of the technology itself, when in fact the focus and attention needs to be on the associated tools that are needed in order to derive maximum benefit from the NLP output – because NLP generates data.
In fact, it would be difficult to argue that healthcare can create the data necessary for the road ahead without NLP. But is generating data enough?

NLP is a technology that generates data; it is not a technology that utilizes information. Therein lies the fundamental disconnect between what is batted about and what is truly needed, and useful. So, what is needed beyond the generation of the data? Actually, to obtain the maximum benefit from the NLP, quite a bit is required.

Enabling technologies to consider

Preprocessing – What did you say? More importantly, What did you mean?

One of the most difficult aspects of natural language processing is the ability for the technology to put what the clinician says into proper context. This exists largely because of the high degree of variability in what physicians say. Physicians are rather adept at saying whatever they feel is necessary to accurately capture the patient’s story.

This is especially challenging on reports that are dictated, ‘the old fashioned way’, because there is no template for the NLP to leverage. Physicians regularly use familiar terms that are no longer compliant with the current Meaningful Use cCDA (Consolidated Clinical Documentation Architecture) standard – terms like ‘Impression’ (cCDA: Assessment) and ‘Lab Data’ or ‘Radiographic Data’ (cCDA: Results).

Is the breast cancer a new or current diagnosis? Was it meant to be a past medical history item? Or perhaps it is a family history item? This ‘tuning’, or what some vendors refer to as ‘learning’, is paramount, because without an adept preprocessor success with NLP is limited and false positives create unintended results – clean up – which equates to inefficiency.

The goal is to efficiently and effectively create usable clinical data from all sources of information, and without a preprocessing mechanism this is extremely difficult, if not impossible, to accomplish.

Rules engine – just as sight does not equal vision; NLP does not equal data-driven.

Rules engines can contemplate limitless logical combinations of data, and then invoke an action, or multiple actions. More sophisticated systems can apply a value to each of the rules to prioritize these actions, bringing substantial incremental value. In practical use, a rules engine is
programmed to do something when the NLP engine generates certain data, or combinations of data. Without a rules engine there is no automated decision-making that occurs as a result of the NLP.

In other words, what good is it to generate data, if you can’t use it? The rules engine is a cornerstone in utilizing data.

**Workflow engine** – Because ‘search’ is inefficient.

Manually searching for different pieces of information is incredibly time consuming and inefficient. Imagine how cumbersome it would be to check your smartphone for text messages all day if it didn’t notify you! How often would you?

Clinical staff do not have the time to be manually searching for important data as it becomes available. Push notifications inform key staff members of significant events based on a set of predetermined rules configured in the rules engine – in other words, the ‘data’ informs the end user of its existence as it becomes available – just like texting.

The ability to create actionable data is pointless if there is no technology in place to act on the data, make it consumable information, and get it in the hands of the intended consumer(s). The rules engine and workflow engine work hand-in-hand to make the NLP data usable information.

**Getting to data-driven**

Data-driven. What does it mean to be data-driven in health care?

RecordsOne defines data-driven as getting information to users when, where and how they need it.

Why is that important? Quality. Tumor Registry. Clinical Research. There may be staff or clinicians throughout the facility, or even in another location, that have interest in an admission depending on the individual, their history, their family history, their combination of diagnoses, meds, allergies, etc. The possibilities are too vast to contemplate, especially for individual staff or clinicians tasked with reading through mounds of documents in the midst of a hectic day.
Why NLP isn’t enough

When Mr. Orange, a 82 YO hypertensive, diabetic gentleman, with CHF & SOB presents, it may be that Quality wants to know why he isn’t on Warfarin, and that Research wants to consider him for the new inhaler Trial starting next month (because of his COPD and long-time Advair use), and Population Health wants to keep tabs on him because of his Diabetes and his potential for readmission.

However, if some of the key information about Mr. Orange is trapped in the free text, most systems will fall short of delivering this important information to the users who need it, even if the programming rules are well written.

In order to provide the best opportunity for success, you need to deploy better tools – tools to generate data and tools to utilize information, even when there is only free text to work with.

NLP alone will not conquer these obstacles. NLP will, at best (depending on the NLP), generate the key clinical data points from the text. But that’s the extent of it – extracting the nuggets of information hidden in the free text is what NLP can do.

NLP will not utilize the data. To truly leverage the benefits of NLP, you need to incorporate all of the surrounding, complementary technologies that maximize the NLP itself and that allow the information to be utilized.

Summary

In summary, NLP generates usable clinical data, which is incredibly important in powering healthcare initiatives; however, NLP does not utilize the information it creates. To maximize the potential of NLP and achieve data-driven workflow, a combination of tightly integrated technologies is needed.
The v6 Platform

v6 generates and utilizes structured clinical data.

v6 generates data when there isn’t any – from unstructured clinical notes – using our best-of-breed Natural Language Processing (NLP) technology, Tomahawk, and best-in-class IMO terminology services.

v6 aggregates the structured information and utilizes the normalized structured data through an unparalleled tool set to deliver relevant information when, where and how it is needed.

v6 is data-driven workflow.

About RecordsOne

RecordsOne is a developer of technology that aids in the creation and use of clinical documentation to improve efficiency throughout the continuum of care. We are focused on delivering affordable, high-value, enabling technologies to healthcare facilities of all sizes. We believe that solutions should be customized to meet your unique needs, and we believe that everyone should be able to afford advanced technologies that benefit both the patient and the bottom line, so we can collectively improve the quality and cost of health care in the U.S.

Our mission is to “Power CARe” — provide innovative and affordable tools that enable Collaboration, Analytics and Reimbursement (CARe) — like no other solution. Health care needs powerful and affordable tools to improve the quality of care as well as improve efficiency and reimbursement.

RecordsOne developed the v6 Platform to radically change the way health care creates and uses clinical information and data. The v6 Platform provides high-value, low-cost & low-risk solutions that power Collaboration, Analytics & Reimbursement — CARe.