

SearchWorks, SHARE-VDE, YEWNO: The Three Faces of Discovery

Philip E. Schreur
Stanford University
November 24, 2017

Slide 1 – Title

Slide 2 – Three Faces of Eve – film clip

Slide 3 – The Three Faces of Eve

- In 1957, the classic mystery, *The Three Faces of Eve* was released.
- According to the International Movie Database: Eve White is a quiet, mousy, unassuming wife and mother who keeps suffering from headaches and occasional black outs. Eventually she is sent to see psychiatrist Dr. Luther, and, while under hypnosis, a whole new personality emerges: the racy, wild, fun-loving Eve Black. Under continued therapy, yet a third personality appears, the relatively stable Jane. This film, based on the true-life case of a multiple personality, chronicles Dr. Luther's attempts to reconcile the three faces of Eve's multiple personalities.
- The movie starred Joanne Woodward as Eve White and Lee J. Cobb as her therapist, Dr. Curtis Luther.
- The movie is famous for being one of the earliest portrayals of multiple personalities in the cinema and won Joanne Woodward the Oscar that year for Best Leading Actress in a Leading Role.
- It's a fascinating glimpse of our ideas about multiple personalities in the 1950s, how the different personalities showed different facets of the core person.

Slide 4 – Split

- In 2016, M. Night Shyamalan produced his version of a multiple personality thriller called, Split.
- In Split, three girls are kidnapped by a man with a diagnosed 23 distinct personalities. They must try to escape before the apparent emergence of a frightful new 24th
- But in Split, we get a very different idea about what it means to have multiple personalities.
- According to the psychiatrist in the film, the personalities are actually fundamentally different people inhabiting the same body, not facets of a single personality.
- These people can have different physical characteristics, different blood pressures, and in one case, one was legally blind and another had perfect vision. It's a very different perspective on Dissociative Identity Disorder.
- So what if anything does any of this have to do with Discovery, you might be thinking?
- As I thought about the three discovery systems I'll be speaking about today, a traditional BlackLight based discovery system called SearchWorks, a linked-data based discovery system called SHARE-VDE, and Yewno, I began to think about what makes them different from each other.
- Are they simply three faces onto our traditional library data, or are they in fact completely different persons with unique characteristics and goals?
- I hope in my time this afternoon to highlight what makes these discovery services fundamentally different from each other, giving most time to Yewno as the newest personality to have appeared.

Slide 5 – SearchWorks Intro

- SearchWorks is the BlackLight based discovery interface developed at Stanford, let's call it Eve White, the more traditional and well understood personality.
- BlackLight is designed to take advantage of the intense complexity of our MARC data, but flexible enough to absorb simplistic data as well.

- It accomplished this by mapping data from various schemas and formats to the same Solr indexes for discovery.
-
- DO SEARCHWORKS DEMO
-
- SearchWorks is a marvelous example of what we can achieve in a system that fully exploits our MARC data.
- We've spent decades developing its possibilities and it's truly powerful as a discovery environment.
- But every personality has its tradeoffs.
- First these traditional discovery systems are closed discovery systems, although they have some links to the outside, their focus is on creating an intense network of internal relationships to explore in Discovery.
- But as these systems grow, a second issue develops.

SLIDE 6 – Solr Indexes

- Here is an example of a definition of one of the Solr indexes, this the one for Title.
-
- **Go over description**
-
- The problem is that over years of development, we add more and more types of data: minimal records, Dublin Core, MODS.
- These data types often have none of the complex coding that our MARC data has, and so cannot map well to the Solr indexes and facets.
- The more varied data types we add, the less consistent and comprehensive the Solr indexes are, and so Discovery becomes decidedly lopsided.
-
- It's a discovery interface with a traditional personality, trying to keep up with the times, but showing the limits of its reach.

Slide 7 – SHARE-VDE

- The next discovery interface I'd like to highlight is the SHARE-Virtual Discovery Environment, or SHARE-VDE.
- SHARE-VDE is being developed by Casalini Libri and is truly revolutionary in the sense that it is based on linked data, and BIBFRAME in particular.
- BIBFRAME is the new linked data communication format whose development is being spearheaded by the Library of Congress to replace MARC.
- BIBFRAME has gone through significant development lately and is now being used by the Library of Congress in a production setting, and in various projects internationally.
- For those of you not familiar with BIBFRAME, let me go over it briefly

Slide 8 – BIBFRAME

- BIBFRAME DEMO
- SHARE-VDE DEMO
-
- As I mentioned, SHARE-VDE is still in the development phase.
- Let's call it the relatively stable but questioning third personality, well, only relatively stable.
- Many core functions are being updated and new features and designs appear with some frequency.
- This initial phase contains a sampling of 200K records from 16 institutions.
- Casalini has converted these MARC records to BIBFRAME 2.0 and reconciled the data against each other.
- The result is a discovery system created around entities, such as work, author, or publisher and based in linked data. But what does that actually mean? The data pool is still quite small and so can't search for Three Faces of Eve, but instead will do my second favorite search: Dante

- DO SHARE-VDE DEMO

Slide 9 - Yewno

- But now let's move on to the last personality to have appeared, Yewno.
- Let's call it the precocious Eve Black, a bit wild and completely non-traditional.
- According to Yewno's homepage:
 - o Using machine-learning and computational linguistics, Yewno's unique technology analyzes high-quality content to extract concepts, and discern patterns and relationships, to make large volumes of information more effectively understood. This core technology drives our product portfolio and our mission to transform information into knowledge. We're here to encourage curiosity and deeper understanding in the world.
- Before I connect to Yewno itself, there's a few things I'd like to emphasize before we move on.
- First, Yewno works with full text:
 - o Years ago, the Head of our library came to me when I was Head of the Metadata Department wondering if I could catalog the internet.
 - o Well I nearly passed out, but the point was he wanted some sort of controlled access to the data that was on the Web.
 - o And it was very clear to me that there was no way possible to catalog the nearly infinite number of resources on the web.
 - o Yewno finally gives us a solution to this problem by relating all this data together in a semantically consistent way.
- And second, Yewno is not dependent on human assigned metadata. It automatically extracts concepts and discerns relationships between them. It is perfect complement both to full text items we could never afford to catalog and also as an additional discovery layer on top of full text items we may have cataloged by reaching down into the texts themselves to draw out concepts our broad metadata topics would never uncover.

Slide 9 Yewno

- DO YEWNO DEMO
- So in summary,
 - o Yewno is a full text discovery interface, it can analyze any amount of full text that you can make available to it: open web resources, full text articles, material from your digital repository , resources into the millions.
 - o Yewno semantically analyzes that full text relating similar concepts across very disparate document types.
 - o Yewno presents a graphical interface of those relationships so that you can explore topics in context.
 - o Yewno allows you to explore and make new associations between topics your searches reveal.
 - o And Yewno can give you access to the full text of the data it uses for analysis.

SLIDE 10 – Conclusion

- So now we've seen our three faces of discovery.
- The first personality, the more traditional and well understood persona, fully exploits the complexity of our bibliographic data:
 - o We've spent decades creating that data, taking great care with its construction, maintenance, and evolution
 - o It's highly attuned to our users' needs and in a miraculous way draws together resources across all language families and centuries
 - o But the connections are brittle, dependent for the most part by text strings, and is limited to the rich, but closed world of library data
- The second personality, relatively stable but questioning, focuses on the shift to entity based discovery, as opposed to record-based discovery:
 - o It allows us to exploit those internal connections implied by our MARC coding but left unrealized by converting them to linkable entities
 - o And by using the same URIs for those entities that the web uses, it allows us to link out to the richness of the Web and bring back data crucial to our patrons
- Our last personality, wild and non-traditional, works in a completely different way

- It focuses on the use of free text as opposed to metadata or linking through identifiers
 - By analyzing large bodies of free text looking for relationships between various elements within them, it allows a user to discover important connections that will never be linked through the use of preassigned metadata or identifiers
-
- Three very different personalities molded by three very different technologies to solve three very different user needs>
 - But why should we try and reconcile them? In many ways these three distinct personalities complement each other, allow us to approach our complex data and research needs from multiple angles>
 - And who knows, even more personalities may develop as technology changes, each bringing its own unique light on information access and discovery
 -
 - And with that, I'd like to ask if there are any questions ...