**Script**

To give some background, current breast implant data are spread out across several sources, such as FDA PMAs, the Global Medical Device Nomenclature (GMDN) database, and the Global Unique Device Identification Database (GUDID). These sources, even GUDID, which is the most comprehensive – it provides the GMDN name and definition – are not complete. For example, none explicitly indicate the profile or surface type.

However, there’s something more that’s missing, and that is the *automatic classification* of breast implants. This is due to the inherent structure of databases; they don’t help “group” specific data. In the context of the BIA-ALCL project, we’re concerned about possible associations with breast implant surface type and t-cell lymphoma, this classification can be immensely helpful when used in conjunction with data mining MDR data and such.

This is where a domain ontology comes into play. Basically, what a domain ontology does is classify entities in a specific domain. You can think of it as a taxonomy; it’s formal and uses a hierarchical structure to describe things – in this case, breast implants and not organisms.

So, our objective for this project was to create a complete ontology to classify breast implants on their features. In the relative short term, we would be using this ontology to support BIA-ALCL related data analysis, such as acting as a backend dictionary for a text mining tool to mine MDR data. In the long run, we will eventually merge with other domain ontologies, such as the Ontology of Adverse Events (OAE). And finally, because we were working with GUDID as the main source of data, we wanted to help improve the accuracy and scope of the database.

We started by gathering data from all our sources: GUDID, the FDA PMA website; we also downloaded product catalogs from the breast implant sponsors – like I said, none of these databases provided device dimensions, so we needed to get that data from somewhere else. After we had the data, we merged all of it, filtered the relevant data, performed data validation, and then outputted it to an Excel file. After that, we used a Python library called Owlready2 to build the ontology from our extracted data. We also used Protégé 5 to help visualize our ontology and perform additional engineering. Finally, we used a reasoner for automatic classification, which gave us our first version of the ontology.

In Figure 2, you can see the initial design pattern of the ontology, with all of the class information we included: manufacturer, shell surface, identifier, shape, profile, fill type, and PMA submission, all connecting to the main class, ‘**breast implant device**’. And for the hierarchy of **breast implant device**, we have FDA approved breast implants, followed by company brand, style (if the company specifies one), and then the individual breast implant. In addition to class information, we also included annotations such as product code, device catalog number, version model number, GUDID link, PMA link, GMDN name, GMDN definition, and device publish dates.

We’ve created 4 defined classes: saline-filled, silicone gel-filled, smooth, and textured breast implant. What the ontology does is allow us to run a reasoner, in this case ELK, and it will classify all the individual breast implants under **breast implant device** into these classes. For example, it classified 906 textured breast implants and 836 smooth breast implants.

The current Breast Implant Ontology is in its preliminary stages. Right now, it extends GUDID information by adding features, such as surface type, creating those relationships between breast implants, and classifying them. In the future, we hope to add more than just FDA approved devices, such as off-market historical implants and maybe expand to worldwide implants. We’ll also focus on adding individual device dimensions, such as height, width, and fill volume. BIO will be released as an open source to researchers, industries, and organizations.