Indigenous Knowledge and Resource Management in Northern Australia

making collective memory with computers

Charles Darwin UNIVERSITY
Innovation in Education

Australian Government
Australian Research Council

Publication Details: This web page was last revised on 4 th April, 2004

Words, Ontologies and Aboriginal databases

Michael Christie. School of Education Charles Darwin University, Darwin, Australia

http://www.cdu.edu.au
Michael.Christie@cdu.edu.au

Words, Ontologies and Aboriginal databases.

Michael Christie
School of Education
Charles Darwin University

Introduction

This paper tells of a group of people working in the increasingly digitised context of teaching and researching Aboriginal languages and cultures in a university context, and in remote Aboriginal communities. The first phase involved the development of digital archives with CDs and a website for university teaching purposes. The second phase takes us into the work of developing digital object management systems which will allow Aboriginal people to control, configure and utilise their digital resources for themselves in their own local contexts. Can digital technologies be developed which enhance rather than inhibit Aboriginal knowledge traditions? How can a group of people build its collective memory and perform its knowledge in a digitised context? And what happens to the work of collaboratively assessing truth claims in such a context? This paper concerns that part of this work which deals with the role of words (or strings of text) and ontologies in developing systems for Aboriginal digital object management.

In the early 1990s,I was invited to set up a tertiary program teaching Yolngu[1] languages and culture at Northern Territory University, later to become Charles Darwin University (CDU). I enlisted the support of Waymamba Gaykamangu, a Gupapuyngu woman and experienced educator. Both of us had had more than twenty years experience teaching Yolngu languages, and we have now been working together at CDU for more than 10 years. Waymamba and I interviewed many Yolngu for their ideas about how we should proceed, and enlisted the support of a group of senior advisers who were consulted in the various Yolngu communities. They offered much support and some basic principles: there are many Yolngu languages, not just one; students need to understand something of the full spectrum of Yolngu culture through the manifold perspectives of the people who own particular languages and associated pieces of land, song, art, ceremony and so forth; they must learn how land is related to languages and totems, and how kin are always related to a network of land totems, languages and ways of performing.

All political and religious arrangements in the Yolngu world must involve people from opposing moieties and thus differing language groups. This opposition is based on the metaphors and embodiments of mothers and their children (yothu yindi) who are always of opposite moieties. The weaving together of different identities is fundamental to Yolngu agreement-making and to religion.

We spent a year before the program started developing sets of materials that contained a wide range of Yolngu representations of Yolngu life, language, history, and culture in a range of languages. We soon began to develop some experience using digital technology to organise these stories and make them available. Our first move was to make a CD library of Yolngu Literature. We obtained signed permission to reproduce about 200 stories, most illustrated, most of which had been collected in the 1970s on audio tape, transcribed, edited and made into books at the Literature Production Centres at Milingimbi, Galiwin'ku, and Yirrkala. We explored a variety of options, and came up with a simple database of Yolngu stories, using FileMakerPro. (Christie 1997) See Figs 1, 2, and 3 for search interfaces, found list, and object display.

[1] Yolngu people live in Northeast Arnhemland in the Northern Territory of Australia

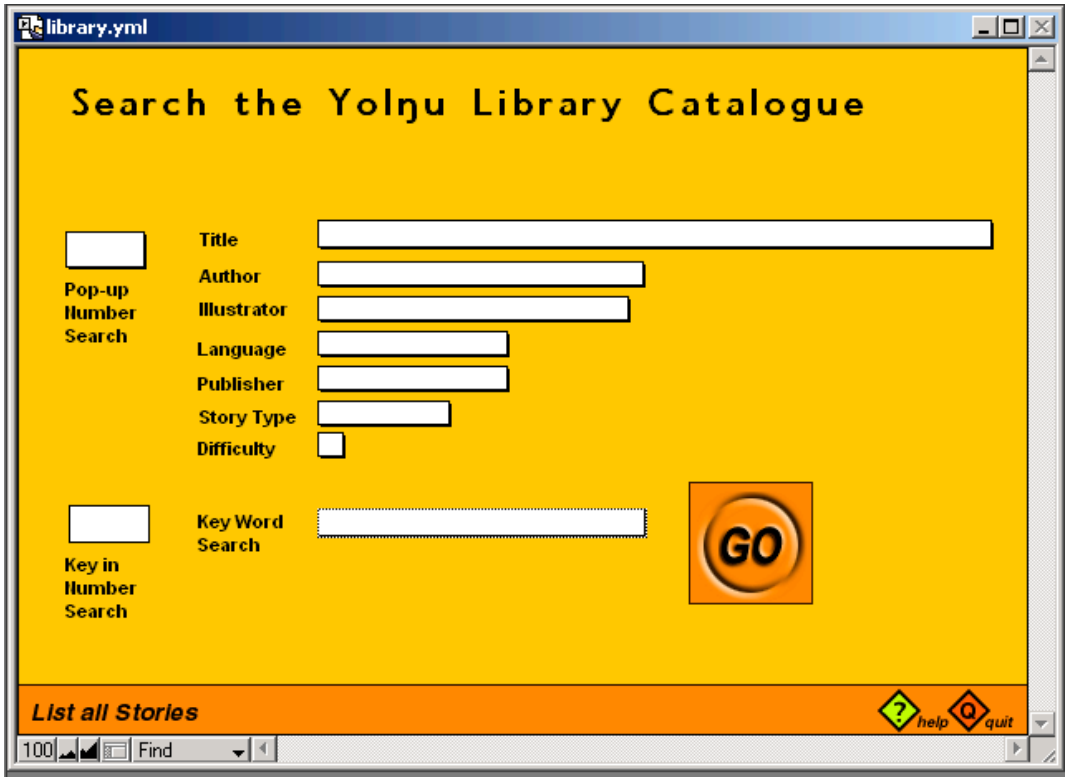

Fig 1: Search interface for CD Library of Yolngu Literature

Fig 2: Results display for CD Library of Yolngu Literature (looking for Djambarrpuyngu language)

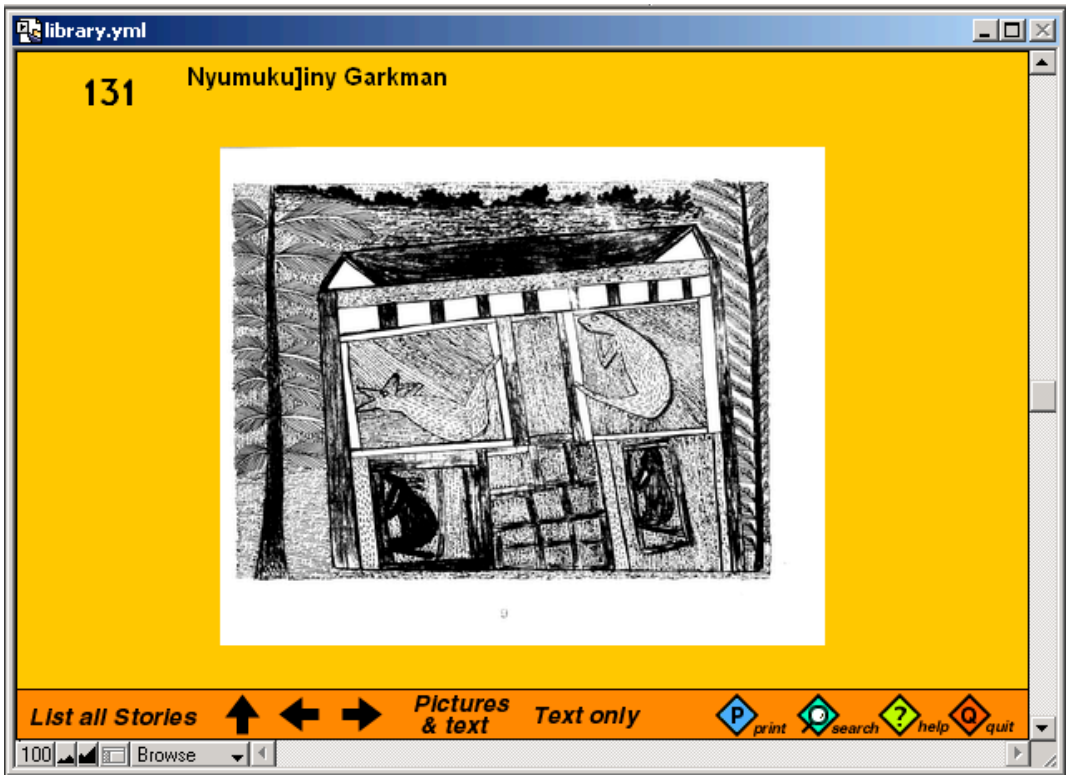

Fig 3: Illustrated page from the book Nyumukuniny garkman, CD Library of Yolngu Literature. (User can also choose text-only)

As soon as the CD Library of Yolngu Literature was finished, more stories came to hand. Instead of starting work towards a second CD, we decided to investigate an online solution. We had already developed a teaching website through a grant from Open Learning Australia (see **www.cdu.edu.au/yolngustudies**). Now, through an internal university grant we received a small amount of funding to develop an online database (see **www.cdu.edu.au/yolngustudies**, click Online Database). We worked with the university's Teaching and Learning Division to set up a database, and provided them with a set of digital resources and a series of metadata text files. The resources were mostly Microsoft Word files, some with pictures, but soon a few sound files, and videos were added. We structured the metadata as far as possible according to the Dublin Core protocols. Together we negotiated suitable search paths and interfaces, and over a few years we ironed out bugs.

Overall, the database, has very usefully served the purposes for which it is designed: making freely available to students and researchers a small range of fully negotiated, authentic representations of Yolngu knowledge in Yolngu languages. The only real frustration has been with the difficulty of uploading from a remote computer. Only a tiny fraction of the stories on the CD found their way to the online database. Our IP agreements with the storytellers and artists (and where deceased, their descendents) were only for the CD not for the Internet. As it has turned out, for a range of reasons, it was found to be useful to have some stories on CD and others on the Internet.

We had by this stage developed an essentially nonAboriginal way of structuring and presenting Aboriginal knowledge. We began to explore some ways in which digital technology can allow Yolngu to represent themselves graphically and vocally, consistent with traditional styles of representation. We were interested in developing ways for digital technologies to assist in the traditional practices of knowledge production, through obviating the hurdles of text literacy. (Christie 2001).

We also became interested in understanding how information architecture both reflects and enacts a politics of knowledge.

> *"Giving something (a story, a video, a photo) a name makes it locatable materially and conceptually. The name provides a textual link to the object describing it to some extent but never exhausting its content. (It may for example identify a storyteller but not anyone else whose presence shaped the telling). The process of naming objects is the beginning of the structuring of knowledge. In the western scientific tradition, the work of naming often assumes a world already objectively structured, and the possibility of a language which "cuts nature at its joints". But (names) are both selective and productive. We too easily assume that the information structures of a database reflect somehow the structures inherent in the natural world." (Christie 2004, p3)*

In Yolngu language a particular word may be a person's name, as well as a place name, or a ceremonial act or object, or a particular connection between people and totems, or land and wind, anything. These links are the inheritance and manifestations of ancestral connections. A "person or group of people could be referred to by a variety of different names in different contexts, and having a mature control over the language meant, in Yolngu terms, being able to select the names to best foreground the meanings significant to the context" (Christie 1994, p26). The way in which the metadata is traditionally sequestered into fields to facilitate searching, actually prevents those flexible political connections between things and names from being real in the database, and gives priority to a western objectivist ontology. For a while we accepted unthinkingly the necessity of metadata fields (title, author, language etc). To some extent we were talked into using them through arguments about interoperability and future-proofing - both of which are now much further down the list of priorities as the everyday work and agendas of Aboriginal people in remote communities are prioritised.

Words, in an Aboriginal language, have their power in the work of knowledge production, by virtue of their potential relatedness to any of the different fields in the metadata structure. In other words they attain their power precisely because they resist that categorization. So whatever the heuristic efficiencies western thinkers may achieve through the distribution of metadata into categories, the same processes may equally inhibit the scientific work of cultures with alternative ontologies. Collapsing the structures of metadata and flattening out their content may enable the creative connecting processes upon which Aboriginal knowledge making depends. The philosophy and practice of the production of indigenous knowledge, actually implies something quite concrete for the coding of software which deals with digital objects in the context of education. Computer ontologies have everything to do with the ontologies at work in the routine practices of everyday life. Fluid ontologies (such as those proposed by Srinivasan & Huang 2004) can serve to break down the barriers between the programmer and the user. In an important sense, the users actually construct the technology both symbolically in their reading of artefacts, and literally in the articulation work through which they produce concrete configurations of artefacts in their everyday practice. (Suchman 2002, p94)

The 'key word search' (Fig 1) did something to allow such connections to be made available because it acted upon all fields. In a sense it allowed them to be flattened. Subsequently, the possibility of discovering links through a text string search was further enhanced on the online database by sending the key-in search string to search both the text data and the metadata, thus flattening out another a priori distinction (between data and metadata). We sought ways to make useful the striking connection

between the Aboriginal and computer ontologies, both of which require strings of text to produce new presences.

Meanwhile we were wrestling with the complexities of Yolngu text (phonological and orthographic) and the difficult task of helping people with varying levels of text literacy in notoriously difficult-to-spell languages, develop a useful text string for database searching. This task was not made any easier any the special font required for Yolngu languages which have characters not found in English (such as the 'tailed n' which represents the ng sound). Examples of the orthography can be seen in Fig 2. (Note that in Fig 3 the story title is represented wrongly because the Yolngu font was not activated for that field so a square bracket is displayed, rather than the underlined n we see in Fig 2.) We wanted to undo the natural tendency of computers to expect people to read and write only English, and of Aboriginal people to expect the same of computers.

The language teaching and associated research at CDU fed the gradual expansion of the Yolngu languages dictionary, a corpus of around 8000 citations to which we are still consistently contributing. We received a small university grant to develop a CD version using FileMakerPro software. This project became our first step in developing friendly and fuzzy search systems. There are two reasons why a fuzzy search was a particularly good idea for these (and other) Aboriginal languages: 1) Yolngu words, because of the phonology and the orthography, can be difficulty to spell. Some phonemic sounds (like different 'd' sounds, or the difference between long and short vowels) are hard to spell for both Balanda (because of pronunciation difficulties) and Yolngu (because of literacy difficulties). 2) In languages with digraphs and difficult vowel systems, alphabetical order is problematic and in a sense arbitrary. Each dictionary maker must decide for him/herself the conventions to be used for alphabetical order. There is a variety of possible solutions (Christie 1993, p3). Fuzzy electronic searching can mitigate both these difficulties.

A fuzzy search depends upon a 'lemma' form which is a simplified searchable form. An algorithm which generates a lemma form for every word, processes the list of citations. The lemma form is a basic approximation of all the complexities of a word. Imagine these words: dhiku, deku, degu, djigu. They each mean something quite different. They all sound rather alike to the nonAboriginal ear, and while of course they all sound quite different to Yolngu, not many Yolngu can pick and peck the correct string on a keyboard(particularly seeing that some letters, like the d, are assigned to keys like '['). In Yolngu languages there are a number of different consonants which sound a bit like an English 'd'. They are written d, d, dh and dj. (In other positions in a word they all can also be confused with t, t, th, and tj.) There is also a vowel 'i' which can be under some circumstances lengthened slightly (phonemically) and spelt 'e'. There is also in some contexts a phonemic distinction between g and k (and in other contexts, no such distinction). Yolngu languages are in fact generally hard to spell largely because in those languages, the orthography and spelling conventions were put into place while the science of field linguistics was in its infancy[2] . From our original list, only one lemma form ('diku') would be generated by algorithm from the citation forms. (The rules in this case would be "for 'dh', and 'd', and 'd' write 'd'", "for 'g' and 'k' write 'g'", "for 'i' and 'e' write 'i'")

[2] If it were done again today, a much better phonological analysis and orthography could be arranged but it is too late.

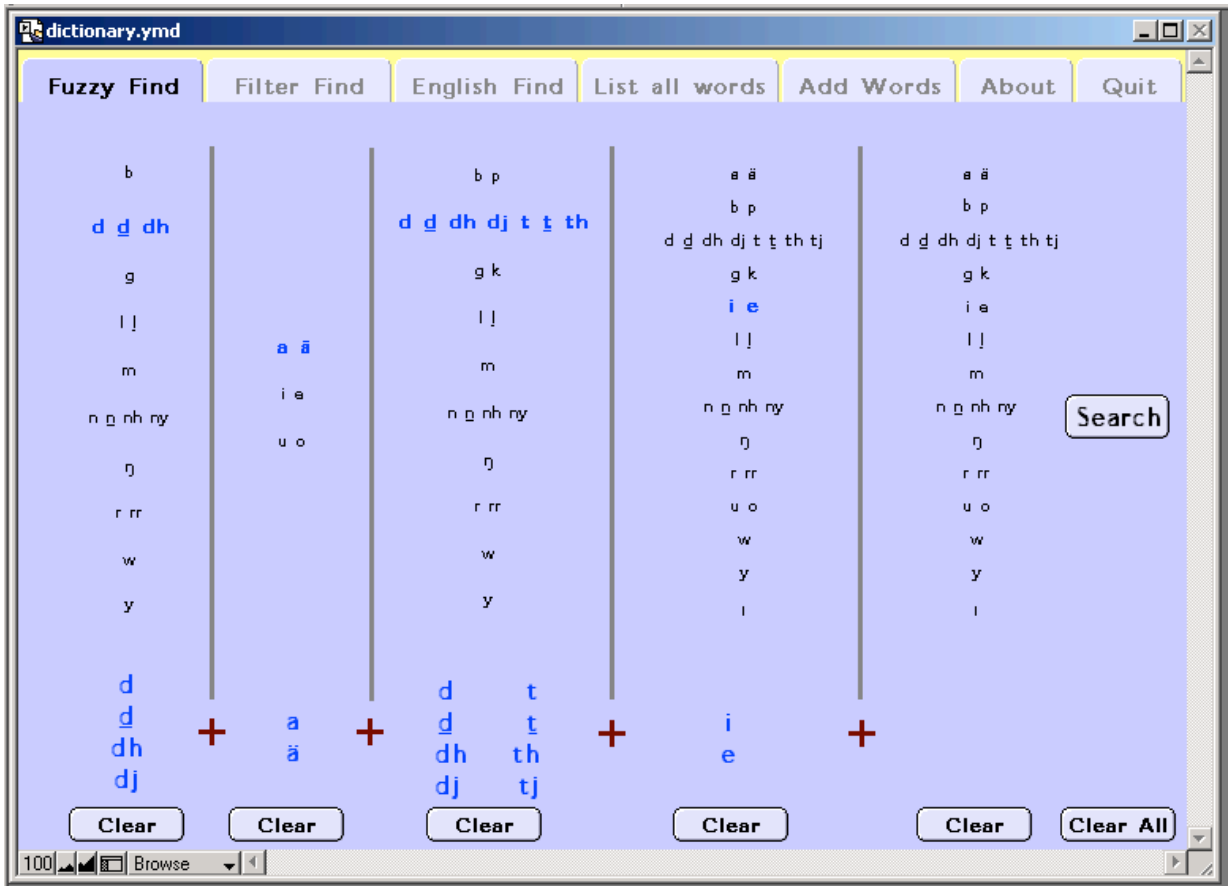

Fig 4: Fuzzy search screen from the CD dictionary of Yolngu Languages.

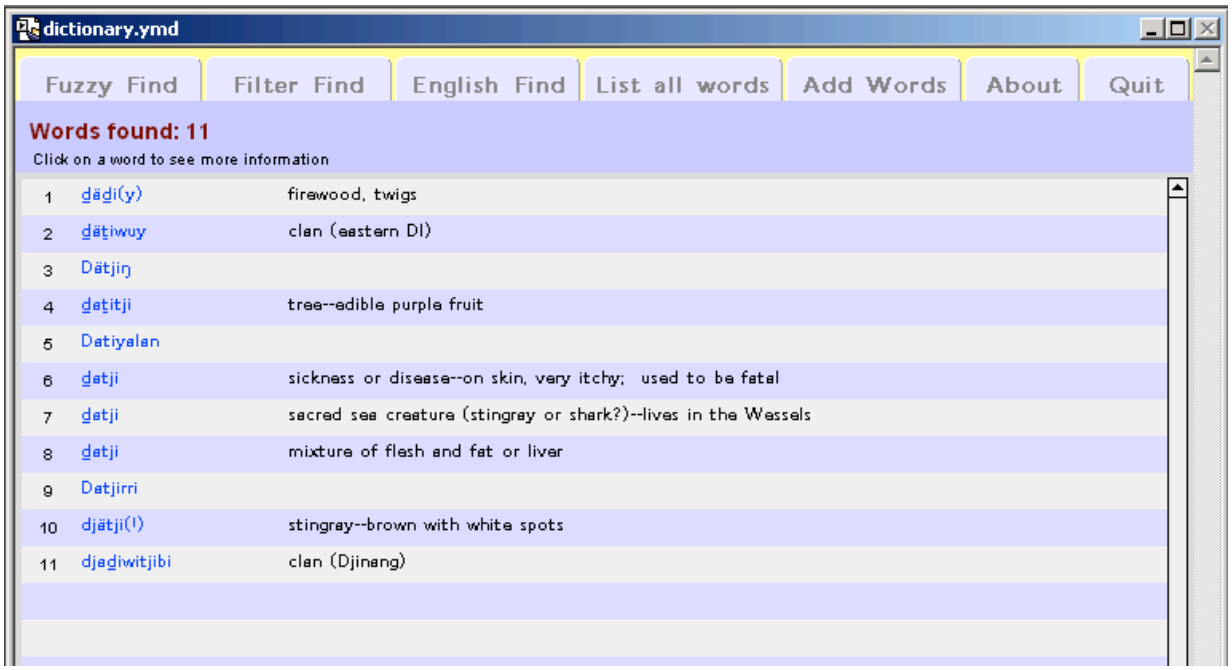

Fig 5: Words gleaned from the combination specified in Fig 4.

Once you have lemma forms available for all words in your data (your dictionary corpus or your database content), you are able to perform a fuzzy search. This can be by voice recognition, keyboard entries, or by clicking on graphics (as in Fig 4). Your fuzzy input will be matched against the lemma list, and from that, generate through elimination, a list of possible words you might be looking for. The fuzzy search screen

from the CD dictionary can be seen in Fig 4. The list of found words generated by the configurations in Fig 4 is displayed in fig 5. (Words without glosses are people's names)

This system is of course designed specifically to respond to spelling difficulties which arise out of the phonological analysis and orthography specific to Yolngu languages. Other systems would need to be developed for other language groups. Some, like prefixing languages for example (Yolngu languages are suffixing) might find such an interface impossible to develop - or come up with a much more useful interface and mechanism for their own cases.

The dictionary also has a 'filter find drop down' search machine (Fig 5) where all the possibilities for the first letter (or digraph) of a word are found in a drop down menu. Choosing one letter (or digraph) refines the list to all words starting with that sound. Then a selection from the possibilities for the second letter leaves us with a list of all words in the glossary starting with these two letters. Then the third, and so on. This is a way of narrowing down the corpus without the use of a lemma form. But you need to be a reasonable speller.

Decisions about how to generate 'smart search' processes like these depend upon the phonological morphological and orthographic conventions in particular languages. In Yolngu languages, the filter find is useful, because digraphs (dh, ny etc) can be offered as units. Clicking on any of the numbered find results opens another page with all the details of that particular word.

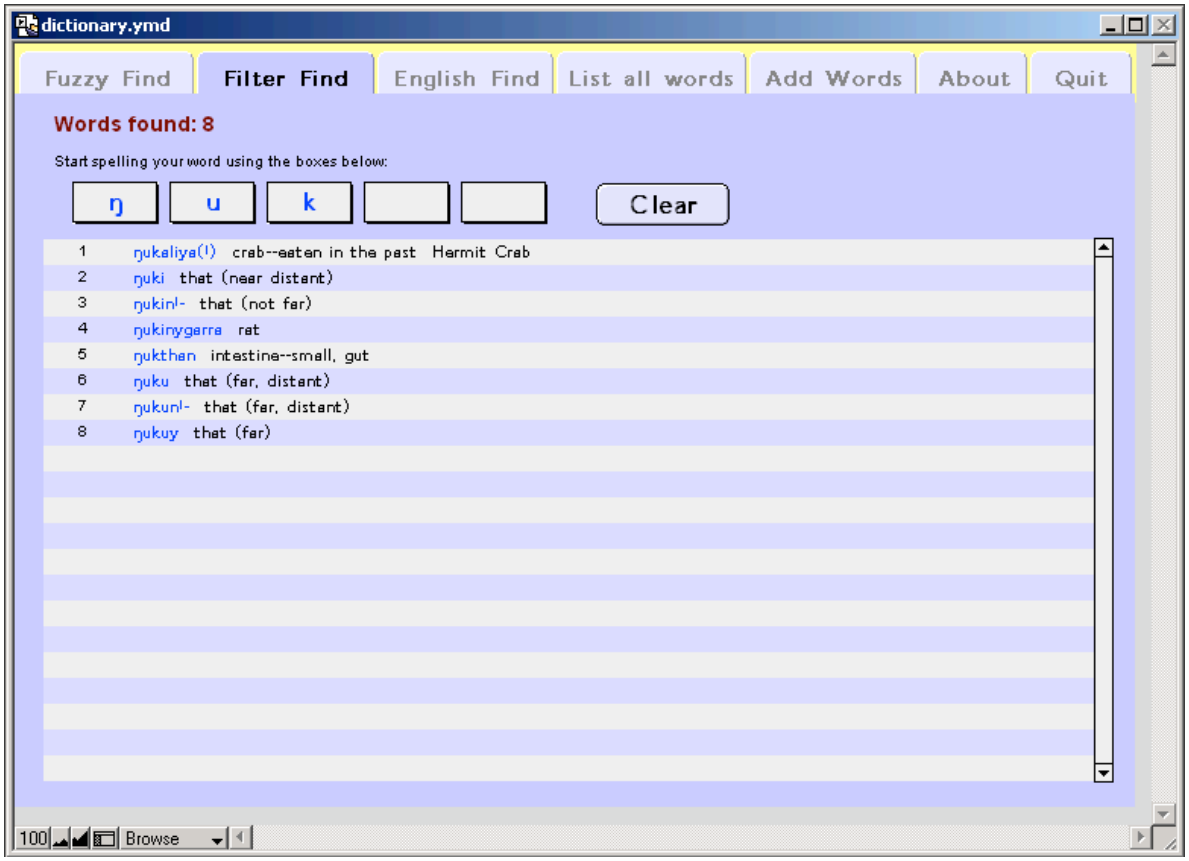

Fig 5: Filter search screen from the CD dictionary of Yolngu Languages.

Phase 2:

The CD library, the online database and the CD dictionary are all, in a sense archives. They are too big for anyone to know all their content so they need metadata structures to aid search. In the past couple of years we have moved into looking at the question

of how digital technologies like those we use at the university can be used by Aboriginal people together in their own contexts in their ongoing work of keeping their traditional ecological knowledge traditions alive. Details of some projects emerging from this work can be found at www.cdu.edu.au/ik. The overall project is looking at ways of making the contents and structures of official archives (and their own digital recording and editing) useful to Aboriginal people on the ground. Local solutions seem to emerge at different points along a continuum which stretches from the official archives at one end, to an ontologically flat and epistemologically innocent database/file management system we have in mind, and which we call called TAMI. TAMI stands for Text, Audio, Movies, Images. It is ontologically flat because it makes no a priori assumptions about the content (or at least only one: it encodes an irreducible distinction between texts, audio files, movies and images). The database ontology makes no attempt to capture the complexity of Yolngu knowledge.

Databases are said to have ontologies insofar as they bear assumptions about the fundamental nature of what they contain. Aboriginal philosophy works through particular ontologies that we hope can be encoded directly into database software. In the past, databases for Aboriginal people have not addressed this issue, and not much has been done to avoid pressing Aboriginal knowledge through the seive of western objectivist ontology.

Data structures have already been identified as influencing the social practices of techno science in the west. Bowker (2000) looking at the relationship between computer databases and biodiversity, demonstrates how different understandings of biodiversity inform and are influenced by the way in which people develop and use databases.

For example, ‘grooving’ is a process whereby the data structures of databases actually affect the way in which we understand the world. Some things in the world are a lot easier to identify or define than others, they make their way without difficulty into databases, and thereby become constitutive of the theory of reality through which we think. Other things however that may be harder to define, or are contested or have fuzzy boundaries, or are radically singular (in that they are unlike anything else), will fall through the cracks. They simply fail to make it into the database. After a while, we develop a represented world within our database, which takes on a particular structure or regularity not so much as a reflection of the reality of the world itself, but much more a function of the data structures that we have chosen to depend upon in the first place.

Bowker and Star (1999) go further, to identify a subsequent process: ‘reverse bootstrapping’ where we start to make assumptions about the nature of the world on the basis of the structures of the data that we are using in our database. In research areas, such as biodiversity and language diversity, fundamental (and often unconscious) assumptions about the nature of reality (for example, the ontological status of an ecosystem or species or language) will inevitably become ‘hard wired’ into the architecture. Their databases go on reflexively to affect the way in which the researchers understand and reinvent the world outside. In the case of Aboriginal knowledge, some things which are perceived to be more charismatic than others, (crocodiles as opposed to algal blooms, for example – both of which are totemically significant in the Yolngu world), find their way into the database, whereas other things, that are equally important in terms of Aboriginal knowledge don’t make it.

There are two directions to search for a solution. One is to encode the complex connectivity of Yolngu knowledge in the data structure. This approach has already been attempted in the “42-level relational database to catch the way the Yolngu people think about the natural world” at the Galiwin’ku Knowledge Centre (*The Australian*, June 10, 2003, pg. 29). The other choice is to do away with the attempt to

hard wire the relationality into the database. To rid it as far as possible of its ontological presumptions, collapse the metadata categories, and create the conditions whereby indigenous owner-users can learn to invoke and encode for themselves the multiple connections which constitute Aboriginal knowledge in the context of database use.

Text-driven searching is still important even when some sort of ontological fluidity, and useful graphic interfaces are available. In the final section of this paper, I describe the functionality of a text-based search which could be applied - with variations - to all the local digital memory solutions which are at work in Aboriginal communities, mixing English and vernaculars, and coping as far as possible, with the linguistic complexities. In each of the contexts we have explored, Aboriginal people are using cameras, recorders, and computers to capture, edit, store, refigure and access digital resources. Most of the time their computer work amounts to file management. The database comes into play only as metadata files need to be generated and attached to objects (or assemblages of objects) in order to facilitate search. We imagine (in TAMI) a database which is first of all, a file management system. Metadata is optional. Text search is optional (and for many will probably be the last resort). TAMI offers a single screen space to store, view, annotate, combine and export digital objects. These local memory databases are possibly unusual (and distinct from archives) in that they may contain a relatively smaller number of objects. The user is the owner of the content. The owner/user has complete control over access to the database. It is ‘my bag of objects’ not ‘everybody’s library’. Where it has metadata they are structured less along the lines of ‘how do I describe this object?’, and more along the lines of ‘How will I find this again when I’m looking for it?’

The plan which we have for the next version of the text search implements a set of principles which tries to develop bridges between community memory, and computer memory. Only under some circumstances, do digital resources need to be located by a text string which is generated by the user. In other circumstances, graphic user interfaces will enable search (as with for example the iLife softwares). Text string searches are only productive if the string which the user generates has a match in the existing (meta)data. The string you are looking for has to pre-exist in the database if your search is to be successful, so we will make only workable strings available to initiate a search. In other words, the user doesn’t generate a string and send it to find a match in the metadata. The user makes a choice from a pre-existing list of words which are already there and uses that to initiate the search. The list of workable strings is generated by a piece of software we call the glossariser, which produces a scrollable, clickable, refinable, and adjustable list of all the text strings (words, numbers and combinations, like file names). This list is always available on screen. Words from the list can be selected and activated as search strings by scrolling and clicking, fuzzy find (as in Fig 4), drop down (Fig 5), key-in (which refines the list in the same way as drop down) or voice recognition. In each of the last four, the search refines the glossary list until the user identifies and clicks a word to initiate the search.

What I have described is a sub-program which can be attached to any database. We are intending to develop it first for the CD Yolngu studies online database, but it should also be useful in a wide range of other contexts.

There is something which Aboriginal metaphysics and computers have in common. Both depend upon a string of text to be productive. Much truth production always already takes place outside of the tyranny of the text string, through art, music, dance, gesture, through performance. However where people choose to use the string of text as a way into digital resource management they need to be able to see what their choices are, and they need to be able to find the best choice without being terrific spellers. That is what we hope to achieve through lemma search capabilities, and

canny ways of helping users locate possibly useful choices in whatever mixture of languages is at work in the glossary.

One way or another, digital technologies are in Aboriginal communities to stay. They can be very useful for traditional knowledge practices, or they can inhibitive and assimilatory. It should be possible for us to bypass the biases of western objectivism and ontology if we use some basic question to inform the design of information technology for indigenous knowledge traditions. What do words achieve in the production of memory and the transmission of knowledge? How do words fit into the social and political context of collective memory? How does information architecture and technologies reflect and produce the possibilities of knowing and of building, sharing and transmitting everyday knowledge?

Bowker, G & Star, S 1999, *Sorting Things Out: Classification and its consequences*, MIT Press, Cambridge Mass.

Bowker, GC 2000, 'Biodiversity, Datadiversity', *Social Studies of Science*, vol. 30, no. 5, pp. 643-83.

Christie, M 1993, *Gupapuyngu English Dictionary*, Yirrkala Literature Production Centre, Yirrkala.

---- 1994, 'Grounded and Ex-centric knowledges: exploring Aboriginal alternatives to western thinking', in J Edwards (ed.), *Thinking: International Interdisciplinary Perspectives*, Hawker Brownlow Education.

---- 1997, 'The Yolngu Literature CD Project', *Ngoondjook*, no. Nov, pp. 31-40.

---- 2001, 'Aboriginal Knowledge on the Internet', *Ngoondjook*, no. 19, pp. 33-51.

---- 2004, 'Computer Databases and Aboriginal Knowledge', *Learning Communities: International Journal of Learning in Social Contexts*, no. 1, pp. 4-12.

Srinivasan, R & Huang, J 2004, *Fluid Ontologies for Digital Muesums*, < http://web.media.mit.edu/~ramesh/publications/FluidOntologies.doc>.

Suchman, L 2002, 'Located accountabilities in technology production', *Scandinavian Journal of Information Systems*, vol. 14, no. 2, pp. 91-105.