



Document QA

Bringing Quality Assurance to the
storage of Knowledge
using a graded approach

The Tyranny of Size

The error rate increases as documents increase in size – a result of too many things to check and the unwillingness to allocate the time necessary to carefully read a large document.

Documents can be grubby at eight pages, and have laughable errors at fifty pages.

The Tyranny of Complexity

Small documents can be complex – mixing many specialties – medical and legal, or electronics and airframes, so no one person really understands it all

Mix in a few more specialties and make the document large, and we will probably have a very poor quality document

Things Simple Document QA Can Find

- Unstable or missing Defined Terms
- Invalid references – products, drawings, standards
- Incoherent Indexed Lists
- Unknown Acronyms
- Erroneous dimensional units

Unstable Defined Terms

An example:

The Contractor

The contractor

The successful tenderer

The Bus Installer

The Bus Installation Company

These are all references to the same entity in a document.

Cutting and pasting from other documents with other definitions of the same thing can make it difficult to work out what is what.

Incoherent Indexed Lists

Indexed Lists can have many faults – a typical one is forgetting what the prologue was:

The Contractor shall:

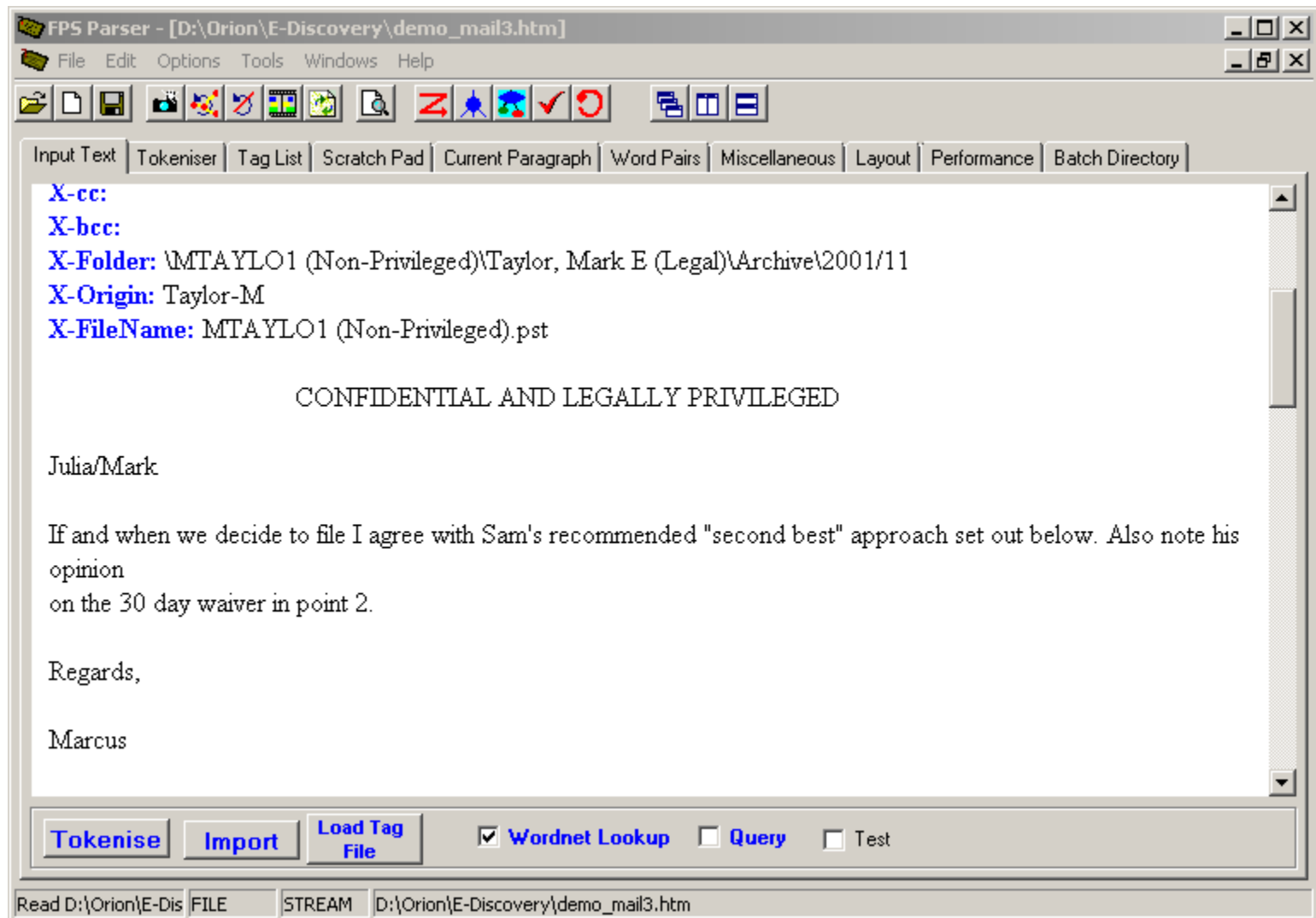
1. Employ best practice ...
2. For OH&S, the Contractor will use...
3. Provide equipment for removal of waste.

One or more of the subitems does not flow from the prologue. This typically happens when the author goes back and edits the document.

Control of Sampling

The QA workload needs to be managed – one way is to sample rarely the documents coming from a clean source, and focus on the dirty sources, or set a threshold on errors before the document is flagged for human intervention.

The QA Tools



The QA structure checking tool can be run manually or in batch mode

Moving to Full Semantics

The next level of QA is full semantic analysis, which can catch many more errors – the type of analysis carried out by a full semantic structure analysis tool like ARIA.

This requires knowledge to be built up in the system, so the full semantic structure can be extracted from the document and automatically analysed in all its complexity.

