

Assembling documents

Two documents were built in memory, saved, reopened in Import mode and merged into this one. Import mode is what permits taking pages out of a document; Modify permits changing them and refuses to let them be extracted. Picking the wrong one is the usual reason an assembly API appears to do nothing.

What was done

Document A	3 pages, a photograph on each, and a link from A1 to A3. 807,069 bytes.
Document B	2 pages, the same photograph on each. 543,650 bytes.
AddPage x 5	Every page of both, copied in. 1,350,160 bytes.
ConsolidateImages	1,043,109 bytes saved - the photograph was embedded twice.
PruneUnusedResources	0 bytes saved.
DuplicatePage(1, 6)	A copy of the first imported page, placed at the end.
MovePage(5, 1)	B2 moved from the end of the run to the front of it.
Split	7 single-page documents, 1,700,825 bytes between them.

What to look for

The pages after this one read B2, A1, A2, A3, B1, A1 - the order the moves above left them in, not the order they were added. The first A1 still carries its link to A3 (1 annotation(s) survived the import); the copy at the end came from DuplicatePage rather than from another import.

Why the two savings differ so much

ConsolidateImages finds XObjects with identical bytes and points every reference at one of them, which pays whenever two merged documents embedded the same logo or photograph. PruneUnusedResources drops what a page names and does not draw with, and a document this library wrote gives each page its own resources - so there is little to find here. It pays on a page imported from a producer that names every font in the document on every page of it, which is why splitting such a document can otherwise give every single-page file the weight of the whole.

Splitting bears that out from the other side: 7 one-page documents come to 1,700,825 bytes against the 307,051 of the document they came from, because each of them carries its own copy of every font and image its page draws with.

B2

Page 2 of document B



A1

Page 1 of document A



[This line links to the last page of this document.](#)

A2

Page 2 of document A



Page 3 of document A



B1

Page 1 of document B



A1

Page 1 of document A



[This line links to the last page of this document.](#)