Document (weighting)	administer	adriamycin	advanced	aids	bleomycin	chemotherapy	clinical	combination	cytotoxic	define	determine	didanosine	dideoxycytidine	dose	doxorubicin	etoposide	kaposi	maximum	oral	orally	patient	pharmacology	relate	response	sarcoma	tolerated	toxicity	treatment	vincristine	weekly	zalcitabine
A (incidence)																															
B (incidence)																															
A (Term Frequency)																															
B (Term Frequency)																															
A (Document Frequency)																															
B (Document Frequency)																															
A (TF-IDF)																															
A (TF-IDF)																															

Using the text of document a and document b below, fill in the modified document-term matrix above using each of the specified vectorization schemes listed. Use the definitions listed below, which may vary slightly from other implementations.

Document A

To determine the toxicity and response to treatment with cytotoxic chemotherapy using doxorubicin (Adriamycin), bleomycin, and vincristine (DBV) for advanced AIDS-related Kaposi's sarcoma in combination with either didanosine (ddI) or zalcitabine (dideoxycytidine; ddC).

Document B

To define the toxicity and maximum-tolerated dose of weekly oral etoposide (VP-16) in patients with AIDS-related Kaposi's sarcoma; to determine the clinical pharmacology of orally administered VP-16 in AIDS patients.

Vectorization Defintions

Incidence $I(w,d)$: Boolean indicator of whether word *w* appears in document *d* or not (1/0)

Term Frequency $tf(w,d)$: Count of appearances of word *w* in document *d*.

Document Frequency $df(w)$: Number of documents *w* appears in.

Term Frequency, inverse document frequency $tfidf(w,d)$: A weighting scheme used to normalize term frequency in a single document by the popularity of that term overall. Defined for our purposes as: $tfidf(w,d) = \ tf(w,d) * (\log_2 \frac{1+total\ number\ of\ documents}{1+df(w)} + 1\ )$