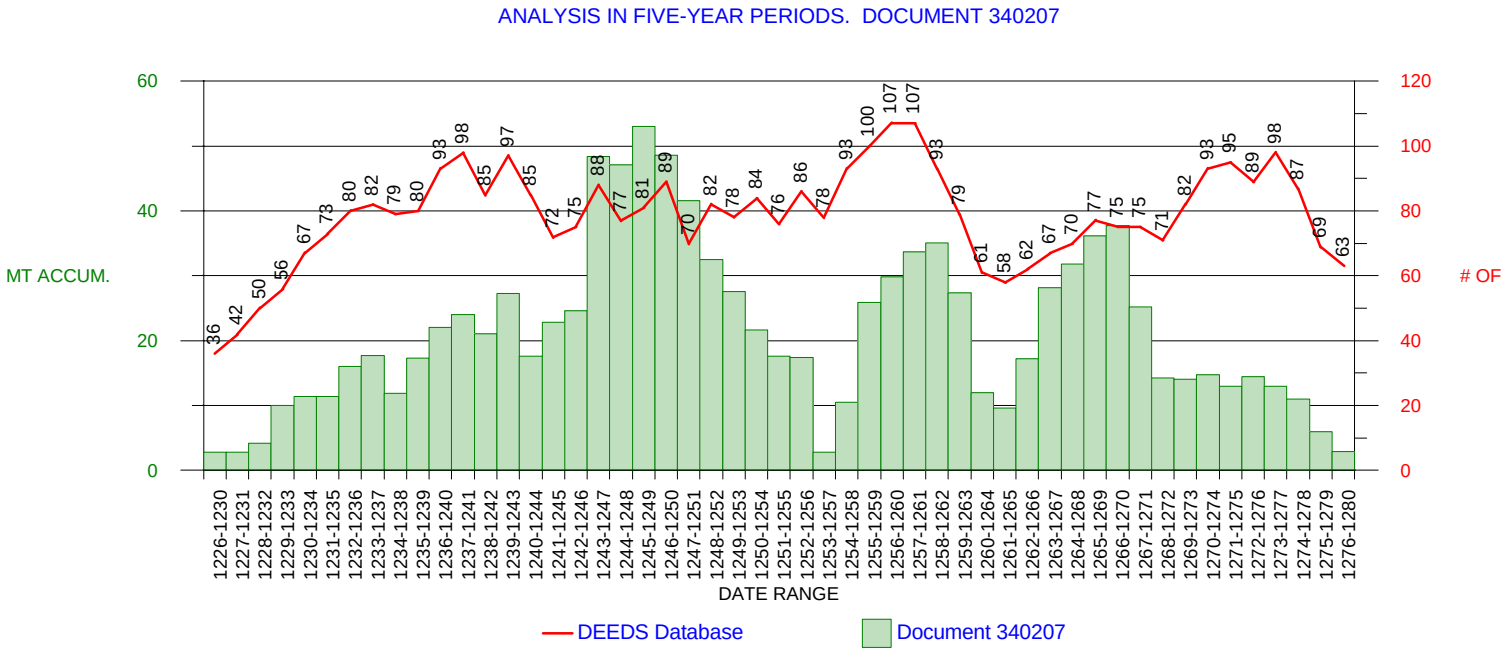
SUPPORTING DOCUMENT

An Example of Word-Pattern Dating in which the Patterns in Charter 340207 (actually dated 1247) are Compared with the 196,632 Patterns in the DEEDS Database.

The calculation of an accurate date is determined by the components of a multiplier (MT), the formula for generating which appears in the appendix on p. 5 of this document.

The graph called "Analysis in five-year periods. Document 340207" shows the distribution of the DEEDS database of documents (red line) over a period of 50 years. The green bars across the graph show the MT numbers accumulated by test document no. 340207 over the same 50-year period when compared with the DEEDS database of dated documents. MT numbers are collected by the test document when some of its word patterns match the ones from the dated documents. MT numbers are assigned to each matching pattern and are generated based on the number of words occurring in it, its lifetime and its currency (frequency of occurrence). Generally, the longer the pattern, the higher the MT number it collects. It should be noted that the MT numbers accumulated by document no. 340207 are not related to the number of dated documents in the DEEDS database, but to the number of word patterns that match the dated documents.



The table entitled "Analysis of the 12 Most Similar Documents" shows the MT number accumulated per document by no. 340207 when compared with entries in the DEEDS database. These 12 documents are the most similar documents to 340207 in terms of word matching. The entries in red indicate documents bearing the same date as no. 340207 and in green those dated within a range of ±2 years of no. 340207. The occurrences of similar documents around the proposed 5-year date-range (1245-1249) for no. 340207 in the previous graph confirm the possibility that no. 340207 was written during that time.

Analysis of the 12 Most Similar Documents

Document	MT Accum.	Year
340203	**13.21**	**1247**
340414	**11.10**	**1247**
340925	**10.75**	**1245**
340063	**10.33**	**1246**
340458	9.89	1252
340857	8.84	1259
340958	8.49	1266
340720	8.00	1267
340708	7.78	1243
340924	**7.02**	**1245**
340692	6.87	1252
340987	**6.61**	**1249**

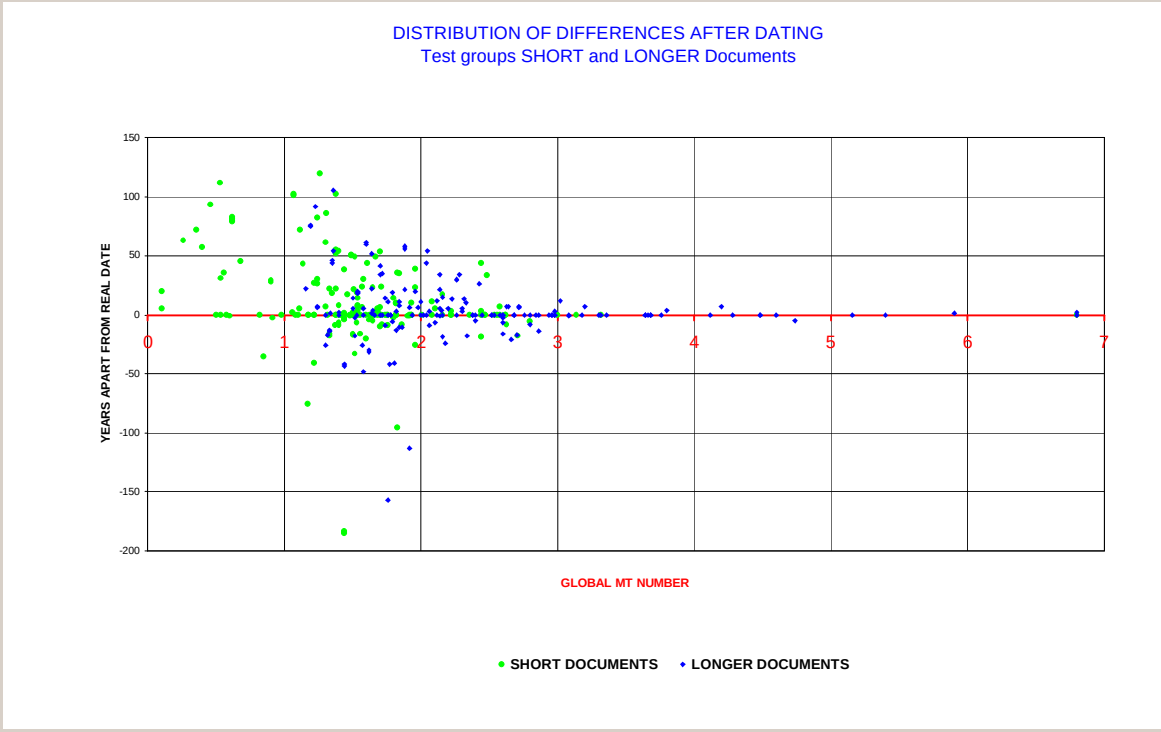
The following page, entitled "Analysis of Currency and Lifetime of Matching Word Patterns" shows the best 30%, i.e. those with the highest MT number, of the patterns in no. 340207 which have counterparts in the DEEDS database of dated documents. Word patterns in blue contain the real date of no. 340207. The Column "Assigned MT" represents the number associated with the matching pattern. Patterns with high currency and short lifetime are qualified with a high MT number and are considered to be the most important for the dating process.

The average of the 10 highest MT numbers (global MT number) gives an indication of the quality of the date approximation. According to experimental data, an MT number greater than, or equal to, 3 indicates an accuracy of ±5 years. Test document 340207 has a global MT number of 3.135.

Position within the document	Length of word-pattern (Words)	Word Pattern	First occur-rence	Last occur-rence	Life-time	Occur-rences	Cur-rency	Assign-ed MT
71	22	ECCLESIE ET CANONICIS BENE ET IN PACE LIBERE ET QUIETE IN LIBERAM PURAM ET PERPETUAM ELEMOSINAM ET EGO IN CUIUS HIIS TESTIBUS	1243	1243	1	1	1.0	4.80
71	16	ECCLESIE ET CANONICIS BENE ET IN PACE LIBERE ET QUIETE IN LIBERAM PURAM ET PERPETUAM ELEMOSINAM	1252	1266	15	3	5.0	3.15
89	16	IN CUIUS HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P LE SEINTER TUNC PREPOSITIS P	1247	1247	1	1	1.0	3.60
56	14	HABUI UEL HABERE POTUI SINE ULLO RETINEMENTO MICHI ET HEREDIBUS MEIS TENENDUM ET HABENDUM	1267	1267	1	1	1.0	3.20
91	14	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P LE SEINTER TUNC PREPOSITIS P	1247	1247	1	1	1.0	3.20
91	13	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P LE SEINTER TUNC PREPOSITIS	1247	1247	1	1	1.0	3.00
71	12	ECCLESIE ET CANONICIS BENE ET IN PACE LIBERE ET QUIETE IN LIBERAM	1246	1252	7	3	2.3	2.80
81	12	IN LIBERAM PURAM ET PERPETUAM ELEMOSINAM ET EGO IN CUIUS HIIS TESTIBUS	1245	1266	22	3	7.3	1.63
56	11	HABUI UEL HABERE POTUI SINE ULLO RETINEMENTO MICHI ET HEREDIBUS MEIS	1234	1259	26	9	2.9	2.60
71	11	ECCLESIE ET CANONICIS BENE ET IN PACE LIBERE ET QUIETE IN	1239	1248	10	2	5.0	2.28
89	11	IN CUIUS HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P	1270	1270	1	1	1.0	2.60
15	10	DEDI DEO ET ECCLESIE SANCTE MARIE DE P ILLUD MESUAGIUM	1268	1268	1	1	1.0	2.40
71	10	ECCLESIE ET CANONICIS BENE ET IN PACE LIBERE ET QUIETE	1239	1259	21	3	7.0	1.50
91	10	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P LE	1247	1247	1	1	1.0	2.40
14	9	MEORUM DEDI DEO ET ECCLESIE SANCTE MARIE DE P	1247	1259	13	2	6.5	1.51
74	9	BENE ET IN PACE LIBERE ET QUIETE IN LIBERAM	1261	1261	1	1	1.0	2.20
89	9	IN CUIUS HIIS TESTIBUS P TUNC MAIORE P AURIFABRO	1269	1270	2	2	1.0	2.20
91	9	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET P	1269	1274	6	2	3.0	2.20
15	8	DEDI DEO ET ECCLESIE SANCTE MARIE DE P	1245	1267	23	7	3.3	2.00
26	8	CUM PERT QUOD EST INTER TERRAM QUE FUIT	1241	1241	1	1	1.0	2.00
79	8	ET QUIETE IN LIBERAM PURAM ET PERPETUAM ELEMOSINAM	1239	1239	1	1	1.0	2.00
81	8	IN LIBERAM PURAM ET PERPETUAM ELEMOSINAM ET EGO	1249	1249	1	1	1.0	2.00
91	8	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO ET	1273	1273	1	1	1.0	2.00
6	7	P PRO SALUTE ANIME MEE ET OMNIUM	1241	1241	1	1	1.0	1.80
28	7	QUOD EST INTER TERRAM QUE FUIT P	1266	1266	1	1	1.0	1.80
29	7	EST INTER TERRAM QUE FUIT P ET	1267	1267	1	1	1.0	1.80
30	7	INTER TERRAM QUE FUIT P ET TERRAM	1240	1240	1	1	1.0	1.80
34	7	P ET TERRAM P FILII P IN	1281	1281	1	1	1.0	1.80
56	7	HABUI UEL HABERE POTUI SINE ULLO RETINEMENTO	1259	1259	1	1	1.0	1.80
87	7	ET EGO IN CUIUS HIIS TESTIBUS P	1246	1246	1	1	1.0	1.80
91	7	HIIS TESTIBUS P TUNC MAIORE P AURIFABRO	1247	1285	39	7	5.6	1.45
6	6	P PRO SALUTE ANIME MEE ET	1229	1272	44	10	4.4	1.42
15	6	DEDI DEO ET ECCLESIE SANCTE MARIE	1233	1233	1	1	1.0	1.60
17	6	ET ECCLESIE SANCTE MARIE DE P	1275	1275	1	1	1.0	1.60
26	6	CUM PERT QUOD EST INTER TERRAM	1240	1240	1	1	1.0	1.60
28	6	QUOD EST INTER TERRAM QUE FUIT	1237	1247	11	3	3.7	1.60
58	6	HABERE POTUI SINE ULLO RETINEMENTO MICHI	1255	1255	1	1	1.0	1.60
87	6	ET EGO IN CUIUS HIIS TESTIBUS	1233	1267	35	7	5.0	1.40
93	6	P TUNC MAIORE P AURIFABRO ET	1250	1261	12	3	4.0	1.60
94	6	TUNC MAIORE P AURIFABRO ET P	1234	1236	3	2	1.5	1.60
99	6	P LE SEINTER TUNC PREPOSITIS P	1247	1247	1	1	1.0	1.60
102	6	TUNC PREPOSITIS P FILIO P ET	1236	1236	1	1	1.0	1.60
6	5	P PRO SALUTE ANIME MEE	1262	1262	1	1	1.0	1.40
10	5	MEE ET OMNIUM PARENTUM MEORUM	1196	1196	1	1	1.0	1.40
16	5	DEO ET ECCLESIE SANCTE MARIE	1229	1233	5	3	1.7	1.40
28	5	QUOD EST INTER TERRAM QUE	1259	1259	1	1	1.0	1.40
31	5	TERRAM QUE FUIT P ET	1190	1190	1	1	1.0	1.40
49	5	EX PARTE AQUILONARI ET QUICQUID	1260	1267	8	2	4.0	1.40
82	5	LIBERAM PURAM ET PERPETUAM ELEMOSINAM	1289	1289	1	1	1.0	1.40
93	5	P TUNC MAIORE P AURIFABRO	1267	1274	8	2	4.0	1.40
94	5	TUNC MAIORE P AURIFABRO ET	1250	1250	1	1	1.0	1.40

The following graph shows the accuracy of the dating process in relation to the global MT number associated with the test document (average of the 10 best MT numbers collected by the test document). The horizontal axis represents the global MT number associated with the test document and the vertical axis represents the difference between the real date and the date obtained through the dating process previously described. The green series belong to a test group of 150 short documents averaging 100 words. The blue series represent a group of longer documents averaging 480 words.

It is obvious that shorter documents will tend to have smaller MT numbers because they contain fewer words for comparison. The blue series of longer documents, on the other hand, includes several documents with an MT number greater than, or equal to, 3.



CONCLUSION

The accuracy of the dating process depends on the test document producing long and good quality word patterns. The first step for dating an undated document, therefore, will be to determine its global MT number.

The procedure for determining the MT number (explained in the appendix on the next page) is under constant revision. The test documents have shown a very consistent pattern of accuracy in the dating process. Documents with a global MT number greater than, or equal to, 3 are considered good candidates to be dated accurately within ±5 years.

The global MT number associated with the test document is also an indicator of the accuracy of the date proposed for it by the dating process. In the graph above, the difference between the real date and the proposed date (Y-axis) increases exponentially as the MT number decreases. That exponential curve describes the accuracy of the dating process as a function of the MT number.

Undated documents unable to produce a high MT number can be dated with less accuracy, but as the process is refined and more dated documents are added to the database, the degree of accuracy will improve.

APPENDIX

FORMULAE FOR GENERATING THE MT NUMBER

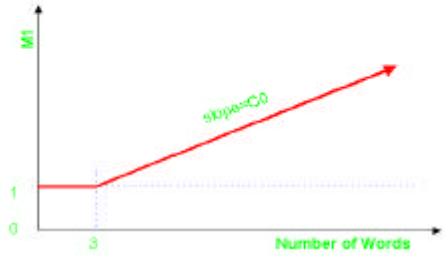
CALCULATION OF MT FOR EACH WORD PATTERN IN A DOCUMENT

I. MT = M1 * M2 * M3

where
MT is Total Multiplier
M1 is coefficient for Number of Words
M2 is coefficient for Lifetime
M3 is coefficient for Currency

II. M1 = 1 if Number of Words <=3
M1 = 1+ C0*(Number of Words - 3) if Number of Words > 3

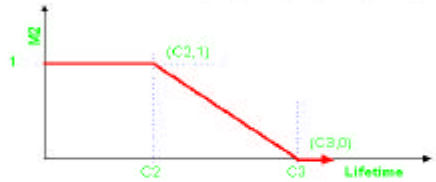
where
C0 is a coefficient in (0.1, 0.2, 0.3, 0.4, 0.5)



III M2 = 1 if 0 <= Lifetime <= C2
M2 = (- Lifetime + C3) / (C3 - C2) if C2 < Lifetime <= C3
M2 = 0 if Lifetime > C3

where

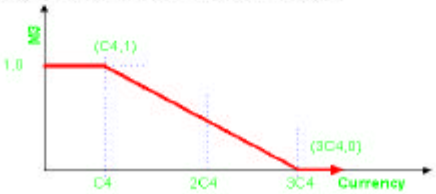
Lifetime = Late Occurrence - Early Occurrence
C1 is a coefficient in (0.5, 0.6, 0.7, 0.8, 0.9, 1.0)
C2 is a coefficient in (10, 20, 30, 40, 50)
C3 is a coefficient in (40, 60, 80, 100, 120)



IV M3 = 1.0 if 0 <= Currency <= C4
M3 = (- 1 / 2 * C4) * Currency + 1.5 if C4 < Currency <= 3C4
M3 = 0 if Currency > 3C4

where

Currency = Lifetime / Number of Occurrences
C4 is a Coefficient in (2, 3, 4, 5, 6)



DOCUMENT 340207

From The Cartulary of Oseney Abbey, vol. 1, ed. H. E. Salter, Oxford, 1929, # 207

Sciant quod ego Agatha filia Agnetis pro salute anime mee et omnium parentum meorum dedi deo et ecclesie sancte Marie de Osen illud mesuagium meum cum pert[inentiis] quod est inter terram que fuit Walteri de Egnesham et terram Thome filii Philippi in parochia sancte Marie Oxonie contra cimiterium ecclesie eiusdem ex parte aquilonari et quicquid in eisdem habui uel habere potui sine ullo retinemento michi et heredibus meis: tenendum et habendum eisdem ecclesie et Canonicis bene et in pace, libere et quiete in liberam, puram et perpetuam elemosinam Et ego [...] In cuius [...] hiis testibus, Nicholao de Stocwelle tunc maiore Oxon, Waltero Aurifabro, et Willelmo le Seinter tunc prepositis, Henrico Simeon, Simone filio Rogeri, Waltero de Osen et aliis