A Description of the Process in which Sample Charter 10662 (actually dated 1274) is Dated by Comparing its Text with the Text of 1,524 Dated Charters in the DEEDS Database.

I. INTRODUCTION

This document is a brief description of the process to date an undated medieval charter by comparing words or word-patterns with other charters for which the date is known. The method used has two critical points: 1) The selection of best word-patterns for dating and 2) the selection of the best date.

The selection of best word-patterns is based on three assumptions: *a) long word-patterns are preferred to short ones, b) word-patterns with a short lifetime are preferred to those with long lifetime, and c) word-patterns which occur often are preferred to those which occur less often*.

The selection of the best date is based on the assumption that there will be a high concentration of word-patterns with high value near to the real date of the charter.

As an example, the following dated charter 10662 was selected from the DEEDS Database to be dated using the word-pattern method.

II. ORIGINAL DOCUMENT TEXT.

The Cartulary of The Knights of St. John of Jerusalem, Secunda Camera, ed. Michael Gervers, New York, 1982, # 662. The text reads:

Omnibus Christi fidelibus ad quos presens scriptum pervenerit Laurencius de Bumstede, salutem in Domino. Noveritis me attornasse Petrum fratrem meum ad tradendam et liberandam seisinam et possessionem tresdecim denariorum et unius oboli, quos Galfridus Gurray michi reddere consuevit de quodam mesuagio quod de me tenuit in villa de Halstede, magistro de Mapeltrestede et fratribus eiusdem domus Deo servientibus; ita scilicet quod habeant et teneant imperpetuum dictum redditum cum omnibus pertinentiis suis de donacione mea sicut in carta a me eisdem confecta plenius testatur. In cuius rei testimonium has litteras meas feci patentes. Datum apud Stouwe die dominica in festo sancti Martini, anno regni regis Edward filii regis Henrici secundo finiente.

III. MODIFICATIONS AND STANDARIZATION.

A process of standarization is carried out in order to avoid break-ups in word-patterns due to numeric expressions, person and place names and minor spelling differences. Punctuation is eliminated, numeric expressions are converted to "#", person and place names are replaced by "P" and spelling differences are eliminated, ie. all "v"s are converted to "u"s.

After the modification and standarization process, charter 10662 reads as follows:

OMNIBUS CHRISTI FIDELIBUS AD QUOS PRESENS SCRIPTUM PERUENERIT SALUTEM IN DOMINO NOUERITIS ME ATTORNASSE P FRATREM MEUM AD TRADENDAM ET LIBERANDAM SEISINAM ET POSSESSIONEM # DENARIORUM ET # OBOLI, QUOS P MICHI REDDERE CONSUEUIT DE QUODAM MESUAGIO QUOD DE ME TENUIT IN UILLA DE P MAGISTRO DE P ET FRATRIBUS EIUSDEM DOMUS DEO SERUIENTIBUS ITA SCILICET QUOD HABEANT ET TENEANT IMPERPETUUM DICTUM REDDITUM CUM OMNIBUS PERTINENTIIS SUIS DE DONACIONE MEA SICUT IN CARTA A ME EISDEM CONFECTA PLENIUS TESTATUR IN CUIUS REI TESTIMONIUM HAS LITTERAS MEAS FECI PATENTES DATUM APUD P DIE DOMINICA IN FESTO SANCTI P ANNO REGNI REGIS P FILII REGIS P # FINIENTE

IV. MATCHING PROCESS.

The result of comparing charter 10662 with the DEEDS database is a collection of word-patterns that are contained in dated charters. Those word-patterns that exist in both charter 10662 and the dated charters are called "matching patterns" and have different characteristics.

Important characteristics of resulting matching patterns are: length (number of words), lifetime (difference, in years, between the last and first occurrence) and currency (the average, in years, between occurrences in the lifetime). Based on those three variables, the resulting matching patterns can be classified in order of importance for the dating process.

Charter 10662 produced 128 matching patterns with lengths from three to fourteen words.

As an example, a resulting matching pattern is shown: **DICTUM REDDITUM CUM OMNIBUS PERTINENTIIS SUIS**. This 6-word pattern is found in a document dated 1275, it has a lifetime of one year and a currency of 1.0 (1 year lifetime / 1 occurrence).

V. WEIGHTING WORD-PATTERNS.

Some word-patterns do not offer much help to the dating process. A word-pattern with a lifetime of 200 years is an example of this, on the other hand, word-patterns with a lifetime of 25 years can be very useful. The first has to be eliminated.

A weighting method is now introduced to facilitate the selection of matching patterns that will be most useful to date the charter. More value wil be given to long patterns, with short lifetime and short currency.

A variable called MT is associated to every matching pattern. MT is defined as follows:

```
MT = M1 * M2 * M3
```

where:

M1 is a coefficient for Length
M2 is a coefficient fot Lifetime
M3 is a coefficient for Currency

M1 is based on the length of the word pattern, and is calculated according to the formula:

```
M1=1                  if Length<=3 words
M1=1+0.2*(Length-3)   if Length >3 words
```

M2 is based on Lifetime (years) and is calculated according to the formula:

```
M2=1                   if  0<=Lifetime<=30
M2=(-Lifetime+80)/50   if 30< Lifetime<=80
M3=0                   if     Lifetime >80
```

M3 is based on Currency (years) and is calculated as follows:

```
M3=1                   if   Currency<= 4
M3=-0.125*Currency+1.5 if 4<Currency<=12
M3=0                   if   Currency >12
```

An MT number can be calculated now for the sample word pattern:

```
M1=1+0.2*(6-3) = 1.6
M2=1.0
M3=1.0
MT=1.6 * 1.0 * 1.0 =  1.6
```

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 10662 has a global MT number of 2.11.

VI. SELECTING WORD-PATTERNS

Once a MT number has been assigned to every word-pattern, a selection of 30% with the highest MT number is made. This percentage is arbitrary and can be modified as experience suggests. Those word-patterns will be the basis for the selection of the most probable date for the undated charter.

The Appendix, entitled "Analysis of Currency and Lifetime of Matching Word Patterns" shows the best 30% of the patterns in no. 10662 which have counterparts in the DEEDS database of dated documents.

VII. ACCUMULATING MT BY PERIOD

After the selection of word-patterns with an associated MT number, it is time to transfer the MT number from the word-patterns to the dates in which they occur.

The sample word-pattern **DICTUM REDDITUM CUM OMNIBUS PERTINENTIIS SUIS** from charter 10662 was found in date 1275. Therefore this date inherits the MT number assigned to it.

In this fashion, year 1275 now has an MT number of 1.60. Year 1275 will increase its MT number if there are more matching word-patterns occurring on that date.

At the end, each date will have as much MT as the sum of all MT numbers inherited from matching word-patterns occurring on that particular date.

Figure 1 shows a period of 60 years with the MT numbers accumulated from the selection of 30% of matching patterns in charter 10662. This graph could be considered a "fingerprint" of document 10662.

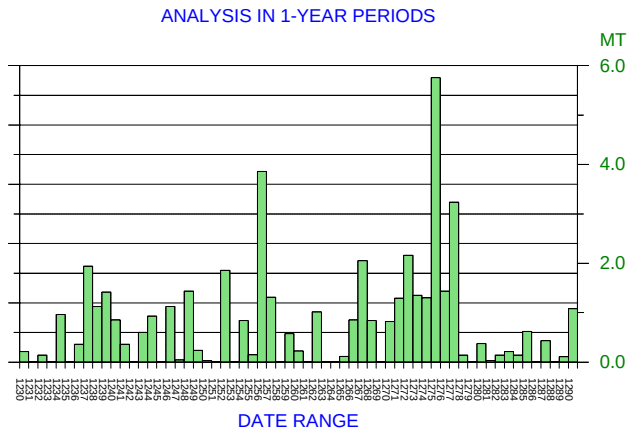

Figure 1. MT accumulated by 10662 per year

Figure 2 shows the accumulation of dated charters (line) and the MT numbers (columns) collected by charter 10662 in 20-year periods. The MT numbers accumulated by charter 10662 are not proportional to the number of dated charters in the DEEDS Database.

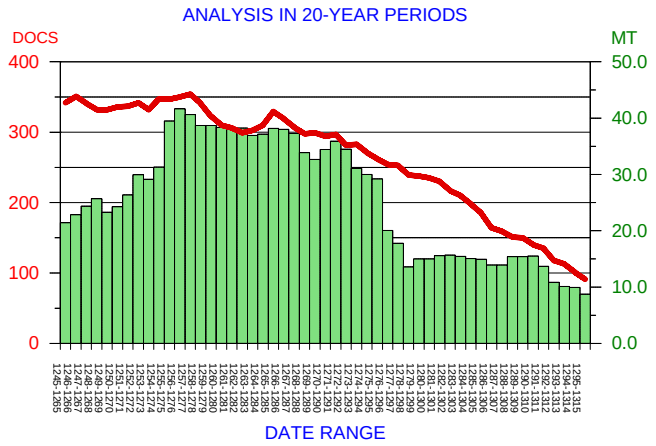

Figure 2. MT(cols.) accum. by 10662 and dated charters in the DEEDS Database (line).

VIII. DATING THE UNDATED CHARTER

Figure 3 represents the ratio MT/document in the same period as Figure 2, accumulating in 20-year periods. The ratio is the highest in range 1262 to 1282. Selecting the period with the highest MT/Document ratio is the starting point in the search for the most probable date for charter 10662.

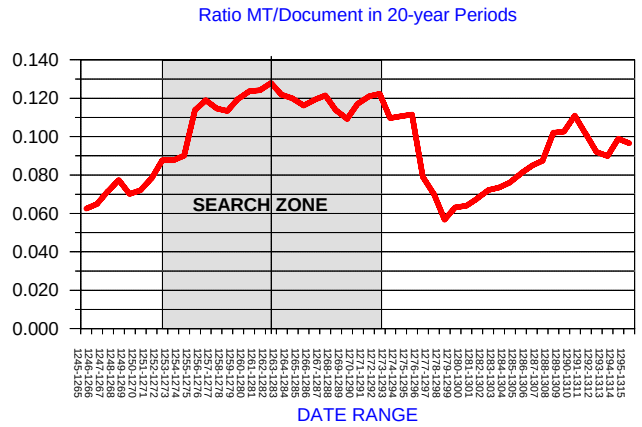

Figure 3. Ratio MT/Document for charter 10662. The search zone is defined ±10 years of peak value.

Once the period with the highest ratio has been selected, it is then expanded by 10 years on each side, generating a "Search Zone" that will contain the most probable date for sample charter 10662. Further steps will narrow the Search Zone down to a 5-year period.

Figure 4 shows the Search Zone with the number of charters and the MT number accumulated in periods of 10 years and in the Search Zone defined previously. Again no proportion is shown between MT and the amount of dated charters.

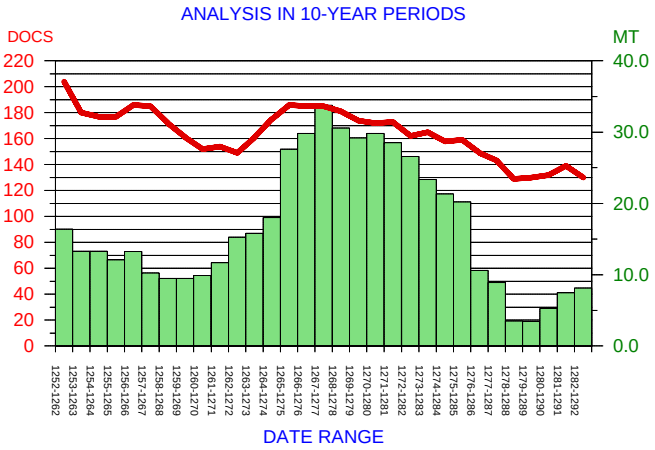

Figure 4. MT (cols.) accumulated by 10662 and dated charters (line) in the DEEDS Database.

Figure 5 shows the ratio MT/Document for the Search Zone with accumulation of 10-year periods. Observe the highest peak. This new peak indicates concentration of word-patterns with high MT numbers.

In order to narrow down the Search Zone, the peak is selected and the zone is defined by expanding 5 years in both directions, as indicated in Figure 5.

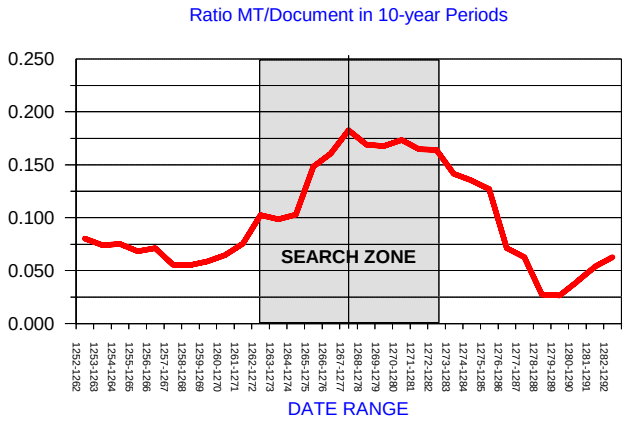

Figure 5. Ratio MT/Document for charter 10662. The search zone is defined ± 5 years of peak value.

The peak corresponding to period 1267-1277 is expanded by 5 years defining a new Search Zone (1262 to 1282). A new accumulation is now necessary to continue the process of reaching a probable date for charter 10662.

Figure 6 shows the MT numbers collected by 10662 (columns) and dated charters (line) accumulated in 5-year periods in the Search Zone (1262-1282). Observe the disproportion of MT numbers and the number of dated charters in period 1271-1275).

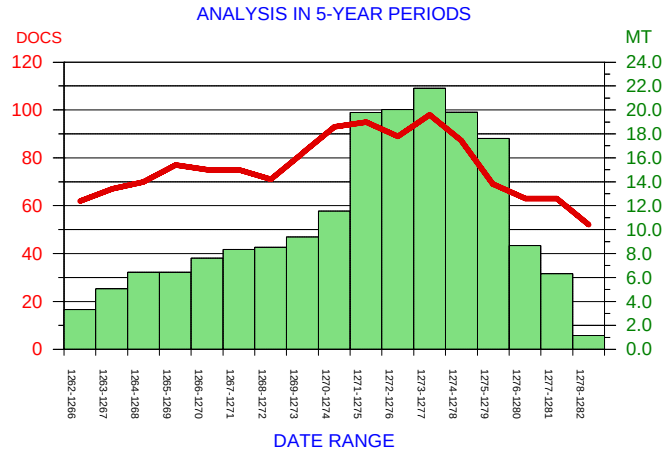

Figure 6. MT (cols.) accumulated by 10662 and dated charters (line) in the DEEDS Database

Figure 7 shows the ratio MT/Document in 5-year periods in Search Zone (1262-1282). Observe the highest peak in the graph (1275-1279), which represents the most probable date for charter 10662.

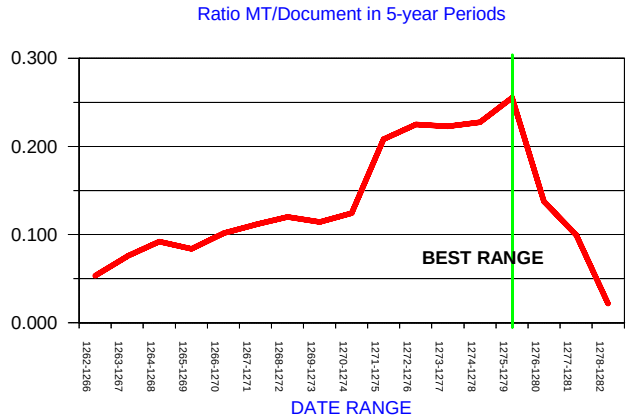

Figure 7. The final selection. The best range is at the peak of the line and corresponds to 1275 to 1279.

Our intention was to describe the dating of a sample dated charter with an accuracy of 5 years. Period 1275 to 1279 is the most probable date for charter 10662.

It is possible to reduce the Search Zone even further to provide an accuracy of less than 5 years and point the dating to a specific year.

IX. CONCLUSION

To test the results, a large selection of dated charters from the DEEDS Database has been dated with this method. The results can be observed in the following table:

Error (years)	Number of Charters	%	Accum. %
0	637	42.9	42.9
1-5	192	12.9	55.8
6-10	130	8.8	64.6
11-20	207	14.0	78.6
21-50	198	13.3	91.9
51-100	104	7.0	98.9
101-168	16	1.1	100.0
TOTAL	1484	100.0	

Accuracy depends on the ability of the undated charter to produce high-valued word-patterns. The global MT number (average of the ten best MT numbers) is the indicator of this ability. Charters with a global MT number greater than 3.0 were successfully dated with an accuracy of ±5 years in 80 percent of cases.

Using statistical methods it is possible to find the equation of accuracy as a function of the global MT number. Preliminary results are already available.

With the accuracy equation, the researcher can easily judge the reliability of each date to be produced after calculating the global MT number.

The size of the control database of dated charters is an influential factor for accuracy. A larger database will allow more word-patterns to be compared, and therefore produce more high-valued word-patterns.

This method has great potential because of its simplicity, cost-effectiveness and accuracy.

APPENDIX

Analysis of Currency and Lifetime of Matching Word Patterns

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
1	14	OMNIBUS CHRISTI FIDELIBUS AD QUOS PRESENS SCRIPTUM PERUENERIT P SALUTEM IN DOMINO NOUERITIS ME	1275	1275	1	1	1.0	3.20
1	12	OMNIBUS CHRISTI FIDELIBUS AD QUOS PRESENS SCRIPTUM PERUENERIT P SALUTEM IN DOMINO	1277	1277	1	1	1.0	2.80
47	9	MAGISTRO DE P ET FRATRIBUS EIUSDEM DOMUS DEO SERUIENTIBUS	1275	1275	1	1	1.0	2.20
97	9	SANCTI P ANNO REGNI REGIS P FILII REGIS P	1308	1308	1	1	1.0	2.20
2	8	CHRISTI FIDELIBUS AD QUOS PRESENS SCRIPTUM PERUENERIT P	1238	1246	9	2	4.5	1.88
96	8	FESTO SANCTI P ANNO REGNI REGIS P FILII	1237	1237	1	1	1.0	2.00
40	7	DE ME TENUIT IN UILLA DE P	1274	1274	1	1	1.0	1.80
81	7	IN CUIUS REI TESTIMONIUM HAS LITTERAS MEAS	1290	1290	1	1	1.0	1.80
1	6	OMNIBUS CHRISTI FIDELIBUS AD QUOS PRESENS	1212	1220	9	2	4.5	1.50
31	6	QUOS P MICHI REDDERE CONSUEUIT DE	1275	1275	1	1	1.0	1.60
40	6	DE ME TENUIT IN UILLA DE	1267	1267	1	1	1.0	1.60
63	6	DICTUM REDDITUM CUM OMNIBUS PERTINENTIIS SUIS	1275	1275	1	1	1.0	1.60
92	6	P DIE DOMINICA IN FESTO SANCTI	1359	1359	1	1	1.0	1.60
95	6	IN FESTO SANCTI P ANNO REGNI	1333	1333	1	1	1.0	1.60
99	6	ANNO REGNI REGIS P FILII REGIS	1239	1239	1	1	1.0	1.60
2	5	CHRISTI FIDELIBUS AD QUOS PRESENS	1218	1218	1	1	1.0	1.40
39	5	QUOD DE ME TENUIT IN	1248	1248	1	1	1.0	1.40
42	5	TENUIT IN UILLA DE P	1254	1254	1	1	1.0	1.40
47	5	MAGISTRO DE P ET FRATRIBUS	1277	1277	1	1	1.0	1.40
95	5	IN FESTO SANCTI P ANNO	1295	1295	1	1	1.0	1.40
97	5	SANCTI P ANNO REGNI REGIS	1256	1256	1	1	1.0	1.40
99	5	ANNO REGNI REGIS P FILII	1234	1270	37	9	4.1	1.19
100	5	REGNI REGIS P FILII REGIS	1273	1273	1	1	1.0	1.40
102	5	P FILII REGIS P #	1256	1256	1	1	1.0	1.40