FORMATS 3
formats

Web and cinema. A comparative analysis of two data bases for on-line research

Lluís Codina and Maria del Valle Palma

e-mail e-mail

1. The invisible Web

As the volume of information available through the WWW increases, the main paradox becomes evident: its very abundance is one of the main stumbling blocks to its consumption.

How can that be? Although it is often said that information is immaterial, the fact is that not only is it not immaterial; its processing is subject to the most material limitations imaginable and, as a result, human capacity for processing information is just as subject to restrictions as any other material process, whether running the 100 metres, writing a poem or cutting down trees.

In the face of the chaotic reality of Internet, on the one hand, and the real restrictions of human cognition for processing information on the other, it comes as no surprise that the (wrongly) so-called invisible Internet or Web, i.e. that part of the Web whose content is inaccessible to the search engines, but which provides the most carefully cross-checked and highest quality information, is increasingly important.

2. Data bases and methodology

In this work we are presenting a sketch for a comparative analysis of the two best cinema data bases to be found on the Web. They are:

- *All Movie Guide* (henceforward AMG)

- *Internet Movie Database* (henceforward IMDB)

Both AMG and IMDB are data bases containing descriptions of films and film-makers structured in the form of entries. Those files provide copious data about both elements (films and film-makers) which can be of great value to scholars.

The data range (see figures 1 to 6) from the most elementary, such as the title and director of each film, to awards and box-office takings, by way of cast lists and technical credits, locations, second version, if any, the slogans used to promote the first release, factual errors or contradictions in the scripts, if any, etc.

Moreover, the entries contain various items of added value information, such as plot synopses, biographies of the film-makers, key words, assessments of the historic and artistic value of the films, etc. Lastly, both AMG and IMDB (especially the latter) have a number of consultation facilities, so that searches can be made from one or more concepts or parameters.

The first aim of this work is to contribute to a better knowledge of two invaluable tools for scholars,

professionals and researchers in the world of the cinema, such as the two data bases mentioned.

The second is to supply criteria for assessing the value and relative quality of those two digital resources and therefore to provide elements to answer questions such as: Is it enough to use one of them or do we need to consult both for a particular piece of research? Do they provide the same information? Are there any differences in the qualitative treatment of the information? What are the qualitative differences between them?

Lastly, we aim to provide some initial elements for a methodology for analysing digital resources of this kind. What we are submitting here is no more than a proposal which, if shown to be useful, would still have to be extended and refined with additional elements. For the time being we shall be content to show some of the parameters which might be included in the methodology and a sample of the results that could be obtained.

For our analysis, therefore, we have compared the descriptive treatment of entries for films and film-makers and the number of documented fields. We have also examined the basic statistics of each data base and carried out performance tests applying four information retrieval tests. Lastly we have applied two qualitative tests to check the relative capacity for providing context or strategic information for each data base.

3. Analysis

In October 2000, AMG (see the figures and tables) supplied information for over 180,000 films and 246,000 film-makers all over the world. Of the two data bases analysed, it is not the one with the largest quantity of information, but it is the one that does the best job of documenting it, i.e., it provides more added value information.

That last point is made particularly clear when we compare the number of documented fields, such as keywords, genre, synopsis and so on in its entries —there are twelve of them— with IMDB, which has only three.

In other words, AMG has twelve fields which do not just contain a simple transcription of data taken from pressbooks or other sources of information; in order to provide the information there has been some kind of intellectual analysis by the contributors and editors. The paradigmatic case are the keywords (subject matter, geographical, chronological and names), plot lines, artistic styles, synopses and criticism. As a result, theoretically there are far more possibilities of carrying out certain kinds of heuristic research in AMG than in IMDB.

Moreover, AMG is outstanding for the quality and efficiency of the presentation of the information. The two qualitative tests, which we have called the Polonsky Test and the Dracula Test, which we will be discussing later, make that quite clear.

For its part, IMDB is outstanding for three things: (1) the sheer amount of information; (2) its splendid information search options and (3) what we might call its complementary information, for example, items about soundtracks or alternative versions of the films (censorship, director’s cuts, television versions, etc.).

IMDB provides information about over 258,000 titles and over 864,000 film-makers. About its search options, we will just say that there are so many different ones that it has created an imaginary IMDB University, from which anyone who has passed the various on-line tutorials required to discover all the possibilities of the data base can graduate.

Lastly, IMDB is a mine of complementary information about the films and the film-makers, from the technical details of the shoot to the advertising slogans used for promotion by way of the budget, a detailed list of the compositions on the soundtrack, related websites, etc.

But even so the area where IMDB tops AMG is the perfect operation of all the options. Whilst on AMG system error messages are frequent, on IMDB all the functions worked perfectly over the two weeks we were carrying out the tests for this article (the first two weeks of October 2000).

3.1. Polonsky Test

The Polonsky Test originated as follows: as everyone knows, when Abraham Polonsky was 38 he directed a film which is regarded as one of the jewels of film noir: *Force of Evil* (1948). However, in spite of its success with audiences and critics alike, Polonsky did not direct another film until 1969 (*Tell Them Willie Boy is Here*). Why would a young director capable of producing a great work as one of his first efforts not direct another for 21 years?

The test consists of seeing whether, with a first review of the data provided by the two data bases, we can: a) discover the existence of that question; b) take a guess at the reason for it. As readers will have deduced, the reason was that Polonsky was put on the notorious Hollywood black list because of his refusal to testify against his colleagues before the Un-American Activities Committee.

If we examine the way in which the two data bases present their searches, we can easily see that only AMG passes the test with flying colours. Just by opening the entry for Polonsky and from line 8 of the biography, hardly having to scroll, we come to the sentence:

Unfortunately, Polonsky was blacklisted in 1950 after refusing to testify before the House UnAmerican Activities Committee. He continued to work as a screenwriter, but didn't get another chance to direct until 1970 (sic) in *Tell Them Willie Boy Is Here.*

It is not that IMDB does not provide that information as well; the point is that because of the highly scattered form in which it does so readers are quite likely to miss it. Specifically, only if they have the patience to reach section 11 of the 14 that make up the entry can they find any such information and even so under the heading “Trivia”, a name which by its very nature will probably dissuade many scholars from reading it in the first place.

That case might be just incidental since, after all, there are not many film-makers with biographical profiles like that one. But we do not regard it as incidental; it is the logical consequence of the rather unfortunate way IMDB has of presenting the information, in contrast with the splendid way in which AMG manages to add and synthesise it. Another qualitative test may well reinforce that idea: the Dracula Test, which we will now explain.

3.2. Dracula Test

The elements on which this test is based are: as everyone knows, for some years Hollywood cinema did not dub the films it produced for distribution abroad; when a film with a certain budget was made, there were two or three casts, one for each language, who shot the same film in the same studios in parallel (that is the basis for the plot of Fernando Trueba’s *La niña de mis ojos*). The corresponding version was then distributed in each language zone: the version with Spanish actors in Latin America and Spain; the French version in France and Canada, etc.

The point is that there is a version of *Dracula* from 1931 which was shot on the same sets (and with the same technicians) where Browning made his highly acclaimed version, but with a team of actors speaking Spanish. That version was directed by George Melford, a veteran director who had made his first film in 1911 and who, when commissioned to shoot this version, had more or less reached his century.

Since the recent recovery of a copy of the so-called “Spanish version” of the film, it has become something of a cult, and among the people who have been lucky enough to see it there is an almost unanimous belief

that it is a film with a personality of its own and as good as, if not better than, Browning's.

The test here worked as follows: let us suppose that a student of horror films decided to find out how many had been shot about the Count Dracula legend. He or she might begin with an obvious option, i.e., to find out which films throughout the history of the cinema have the word 'Dracula' in the title. In that case the information obtained from AMG and IMDB would not be the same. In AMG the film could not pass unnoticed, since on the list of answers to the question, of the 42 titles the film that appears at number 8 reads: '4.5 1931 Dracula (Spanish Version)' (4.5, indeed, is the score out of a maximum of 5).

However, if we ask the same question on IMDB, there is nothing in the answer to attract the scholar's attention. On the other hand, the number of films is much higher: 94 on IMDB. But none of them is Melford's Spanish version. So does the film not appear in the IMDB data base? Yes, but we can only find the file if we know what we are looking for. Thanks to its superior search functions, we can ask for all films that have the word 'Dracula' and which, at the same time, are in Spanish. Then, indeed, the mysterious title 'Dracula/II' appears. If we open the file, we need a little more patience and we have to move through the structure of the entry to reach a commentary that clears up the whole question.

Once again we have seen how the same critical or context information can be found in both data bases, but in AMG it usually "catches the eye" whilst in IMDB it can only emerge through questions it already knows or a really patient reader.

4. Conclusions I

The best data base would be a combination of the two: the efficiency of presentation of the information in AMG and the technical and functional efficiency and amount of information in IMDB.

That means that although for most information needs IMDB is enough, if we want to be sure of obtaining the relevant items in each case, we should also consult AMG.

Moreover, although IMDB may provide more information and more varieties of it, it is no less true that AMG presents it so efficiently that when time for obtaining it is limited it is better to use AMG for that reason.

Nevertheless, the global superiority of IMDB is clear. In our methodology, out of a theoretical total of 24 points AMG scored 12, while IMDB scored 20. In percentage terms, whilst AMG has 50% of the ideal of the 24 points, IMDB has 87%, i.e., 37 percentage points ahead of its rival.

AMG has come very badly out of that scoring because of its constant failures and error messages, a section we have assessed under the name of Technical Performance and in which we have had to give AMG zero, both because of its sparse search functions and the relative lack of complementary information in comparison with IMDB.

Naturally, those results come from out methodology and our measurement system. Another form of measurement and choice of what is to be measured would give different ones.

5. Conclusions II

We have to add some conclusions II, this time of the *information policy* type. It is highly significant, and highly predictable, but none the less insufferable that there is no project in any of our home spheres (Catalonia, Spain, Europe) which can be compared to AMG or IMDB.

There is an urgent need to launch a viable project to provide a similar digital resource, but with two differences: (1), done from out perspective and, (2) defending our interests against those —legitimate but

different— of Hollywood.

We are aware that our demand bears some resemblance to the mouse that suggested belling the cat. It is true that it is easier to say things than to do them, but at least let us say them.

6. Sources

AllMovie Guide
http://allmovie.com

Internet Movie Database
http://www.imdb.com

Internet Invisible
http://www.internetinvisible.com

Invisible Web
http://www.invisibleweb.com

Appendix. Comparative analysis *AMG/IMDB*: Figures and Tables

Figure 1. AMG film entry model

- 01. Title
- 02. Year
- 03. Country
- 04. Running time
- 05. Color/B&W
- 06. Feature/Documentary
- 07. AKA
- 08. AMG Rating
- 09. Director
- 10. MPAA Rating
- 11. Genre/Type
- 12. Artistic/Production Styles
- 13. Flags (moral classification)
- 14. Keywords
- 15. From book by, From story by, From play by, From idea by
- 16. From book, From play
- 17. Set In
- 18. Key name
- 19. Release
- 20. Box office
- 21. Movie budget
- 22. Color type
- 23. Sound by
- 24. Cinematic process
- 25. Production Styles
- 26. Produced by
- 27. Release
- 28. Released by
- 29. See also
- 30. TV/Cable/DSS Schedule
- 31. Preview

- 32. Plot Synopsis
- 33. Review
- 34. Cast
- 35. Production Credits
- 36. Awards
- 37. Related movies
- 38. Movies with same personnel
- 39. Movies on the same theme

Figure 2. AMG film-maker entry model

- 01. Name
- 02. Birth name
- 03. AKA
- 04. Occupation
- 05. Birth
- 06. Death
- 07. Occupations
- 08. Years active
- 09. Countries
- 10. Genres
- 11. See Also
- 12. Biography
- 13. Worked with
- 14. Filmography
- 15. TV Schedule
- 16. Awards
- 17. Birth Name

Figure 3. IMDB film entry model

- 01. Title
- 02. Year
- 03. Director
- 04. Writing credits
- 05. Genre
- 06. Tagline
- 07. Plot Outline
- 08. User Comments
- 09. User Rating
- 10. Cast overview
- 11. Also Known As
- 12. Runtime
- 13. Country
- 14. Language
- 15. Color/B&W
- 16. Sound Mix
- 17. Certification
- 18. Reviews
- 19. Awards and nominations
- 20. Plot keywords
- 21. Memorable quotes
- 22. Trivia
- 23. Goofs
- 24. Soundtrack listing

- 25. Alternate versions
- 26. Movie connections
- 27. Box office and business
- 28. Filming locations
- 29. Technical specs
- 30. Literature listings
- 31. News articles
- 32. Trailers
- 33. Posters
- 34. Photographs

Figure 4. IMDB film-maker entry model

- 01. Name
- 02. Date of birth (location)
- 03. Date of death (details)
- 04. Biography
- 05. Filmography
- 06. Awards and nominations
- 07. Other works
- 08. Publicity
- 09. Agent
- 10. Photo gallery
- 11. News articles
- 12. Official site

Figure 5. Partial view of an **AMG** film entry: Force of Evil, directed by Abraham Polonsky

Figure 6. Partial view of an **AMG** film-maker entry: Abraham Polonsky

Figure 7. Partial view of an **IMDB** film entry: Force of Evil, directed by Abraham Polonsky

Figure 8. Partial view of an **IMDB** film-maker entry: Abraham Polonsky

Table 1. Basic statistics

Data base	*Films*	*Film-makers*
AMG	182,441	246,188
IMDB	258,337	864,419

Source: AMG, IMDB, October 2000

Table 2. Quantitative test results

DB	*Test 1*	*Test 2*	*Test 3*	*Test 4*
AMG	-	23	55	8

IMDB	52	27	79	10

Notes:

Test1 = Films with the word 'Barcelona' in the title

Test2 = Films based on works by Hemingway

Test3 = Films with scripts by Rafael Azcona

Test4 = Films by Luis Buñuel in association with Luis Alcoriza

The symbol "-" indicates that it failed the test

Table 3. Qualitative test results

Data base	***Polonsky Test***	***Dracula Test***
AMG	**	**
IMDB	*	*

Notes:

**= Solved in best way *= Not solved in best way

Table 4. Description and documentary treatment

DB	***NF Films***	***NF Film-makers***	***NDF***
AMG	39	16	12
IMDB	34	12	3

Notes:

NF Films = Number of fields in films

NF Film-makers = Number of fields in film-makers

NDF = Number of documented fields (between the two entry models)

Table 5. Global assessment

DB	***qLt***	***qTi***	***SF***	***PI***	***CF***	***TP***	**Global Assessment**

AMG	4	2	1	4	1	0	**12** (50%)
IMDB	2	4	4	2	4	4	**20** (87%)

Notes:

i) Each section can be assessed between 0 and 4

ii) Meaning of the columns:

qLt = Quality of information

qTi = Quantity of information

SF = Search functions

PI = Presentation of information

CF = Complementary functions

TP = Technical performance

GA = Global assessment, out of 24, absolute and as a percentage