

# SCREENPLAY ANALYTICS

*Stewart McKie*

17 January 2014

---

Screenplay analytics is defined as the use of software tools to analyze and visualize the datafied content of a screenplay in order help improve the quality and/or commercial potential of the script. Screenplay analytics is not about analyzing movies, nor is it about intuitive 'human' analysis of screenplays by reading screenplay documents, but is focused on deriving analytical insight automatically from the evidence of the data.

---

## *Abstract*

In this 3-part article I will:

1. Frame the domain of screenplay analytics
2. Discuss some existing archetypes for screenplay analytics
3. Explore screenplay analytics using my online tool ScriptFAQ

You can find ScriptFAQ at <http://scriptfaq.tripos.biz>.

---

## *About the Author*

Stewart McKie has an MA in Screenwriting and recently completed a PhD at Royal Holloway, London focused on screenwriting 2.0 technology and screenplay analytics. You can find out more on his blog at <http://phd.tripos.biz>.

## Contents:

- [1. Framing Screenplay Analytics](#)
  - [1.1 Digitization vs. Datafication](#)
  - [1.2 Data vs. Metadata](#)
  - [1.3 Analysis and Visualization](#)
  - [1.4 Screenplay Analytic Tools](#)
- [2. Screenplay Analytics: Analysis and Visualization Archetypes](#)
  - [2.1 Circular Archetype](#)
  - [2.2 Linear Archetype](#)
  - [2.3 Nodal or Network Archetype](#)
  - [2.4 Rollercoaster Archetype](#)
- [3. Screenplay Analytics with ScriptFAQ](#)
  - [3.1 ScriptFAQ Script Home](#)
  - [3.2 ScriptFAQ Script FAQ List](#)
  - [3.3 ScriptFAQ Scene FAQ](#)
  - [3.4 ScriptFAQ Role FAQ](#)
  - [3.5 ScriptFAQ Location FAQ](#)
  - [3.6 ScriptFAQ Words FAQ](#)
  - [3.7 Sentiment Analysis](#)
  - [3.8 Analysis Using Other 'Big Data' Tools](#)
- [Conclusion](#)
- [References](#)

# 1. Framing Screenplay Analytics

Analysis of screenplays is not new.

In the first half of the 20th century, the discourse was essentially focused on suggesting the 'rules' that identify 'good' scripts or those likely to have 'commercial' potential. Two early classics in the literature include *The Palmer Plan Handbook* (1919) and Vale's *The Technique of Screenplay Writing* (1944). Lately there has been more emphasis on analyzing screenplays to identify potential box-office hits such as the work by Elishaberg, Hui and Zhang at Wharton or to identify potential Oscar winners by analyzing the 'sentiment' of social media by Gloor at MIT. Many works relating to narrative analysis, while not necessarily specifically focused on screenplay narratives, are also of relevance, such as Chatman's *Story and Discourse: Narrative Structure in Fiction and Film* (1978).

Screenplay analysis really took off after Field re-introduced the Aristotlean 3-Act structure in 1979. The 1980's and 90's saw the rise of a new generation of competitive and commercial screenwriting 'gurus' who published hundreds of screenwriting 'how-to' manuals aimed at 'wannabee' and 'professional' screenwriters that generally used close reading and analysis of a wide range of screenplays to suggest better ways to write and sell movie screenplays. But these works of screenplay analysis are based on reading scripts (as documents) and applying intuitive analysis to the text and not on programmatically analyzing screenplay content as data and deriving analysis from the evidence of the data.

## 1.1 Digitization vs. Datafication

The prevailing mindseye image of a script is that of a dog-eared paper document, bound by 3 brass brads, formatted to specific industry conventions and written in a specific 'legacy' typewriter font: Courier 12. Although paper scripts (screenplay-as-document) still prevail in the movie-making industry they are often communicated, shared and read online in a digital format, such as Adobe's popular PDF. Here, the text of the screenplay has simply been digitized for display and navigation on a computing device screen.

But the digitization of screenplay content is not the same as what Schönberger and Cukier call 'datafication' of screenplay content. It does not 'datafy' the screenplay.

Datafication is the organizing and storage of data in a specific format and medium to facilitate the analysis of the data using software tools (screenplay-as-data). In terms of screenplays this means disassembling the screenplay's textual content into a series of data items that are typically no longer stored in a digital file (like a PDF document) but within some kind of database or at least a file with an identifiable analytic structure such as an Extensible Markup Language (XML) file.

It's quite possible that if a screenplay is datafied every individual word in the textual content is stored as a separate data item and many potential relationships between these words also stored in some way. Note that this changes the granularity of the screenplay text from the 'coarse' level of a script document or file to the 'fine' level of a word or data item. Also that datafication is not intended for human consumption of screenplays (i.e. script reading and intuitive analysis) but for machine consumption (i.e. data querying and evidential analysis).

## 1.2 Data vs. Metadata

When a screenplay is datafied, it is actually two kinds of data that is being created:

1. The data
2. The metadata i.e. data that describes the data

Screenplays comprise both data and metadata. The data is only the spoken dialog and the action/scene description. Everything else is metadata because it provides essential context for the data. Take this simple dialog 'snippet' from *Casablanca* (1942):

ILSA  
Play it, Sam.

'Play it, Sam' is data and ILSA is metadata because it tells us who speaks the data. The other examples of metadata in a screenplay are:

- scene headings that tell us about the context for the scene data
- voice and parenthesis that tells us about how dialog is spoken
- transitions that tell us about the flow of the narrative
- captions that tell us something specific about the context of the data

Note that the metadata associated with a screenplay document is 'sparse' i.e. there is not much of it to assist with data analysis. A datafied screenplay will most likely provide the ability to add significantly more metadata to the screenplay, both to the data and to the existing metadata. For example the character metadata for ILSA could be extended to include:

- her gender and age
- is she a protagonist or antagonist?
- images of potential casting choices for the part in the movie

A rich set of metadata is an important element in extending the potential of screenplay analytics.

## 1.3 Analysis and Visualization

Screenplay analytics is about using software tools to analyze and visualize datafied screenplays.

Specific analysis techniques are used to return datasets from the screenplay data in order that this dataset can then be visualized in some way to assist in understanding, exploring or 'asking questions of', and 'gaining insight' from the dataset. There are a number of analysis techniques that can be applied to screenplay data in much the same way that they are applied to many other datasets, such as entity analysis, predictive analysis and sentiment analysis for example. And there are also a number of visualization formats that can be applied to screenplay data such as tables and charts and infographics for example.

As an activity domain, screenplay analytics overlaps with other kinds of analytic spaces including text and content analytics, which, as their names suggest, are focused on the analysis of what is typically termed 'unstructured' data as found in digital documents or blog posts or tweets for example.

## 1.4 Screenplay Analytic Tools

There are already many 'generic' analytic tools available to use to analyze screenplays even if the tool is not specifically intended to analyze screenplay data and the screenplay content itself is not strictly 'datafied'. For example to create a word frequency cloud from a screenplay you can simply paste the screenplay text into one of a number of word cloud generators (e.g. [wordle.com](http://wordle.com) or my own [scriptcloud.com](http://scriptcloud.com)) and get cloud visualization of the most frequent words in the script.



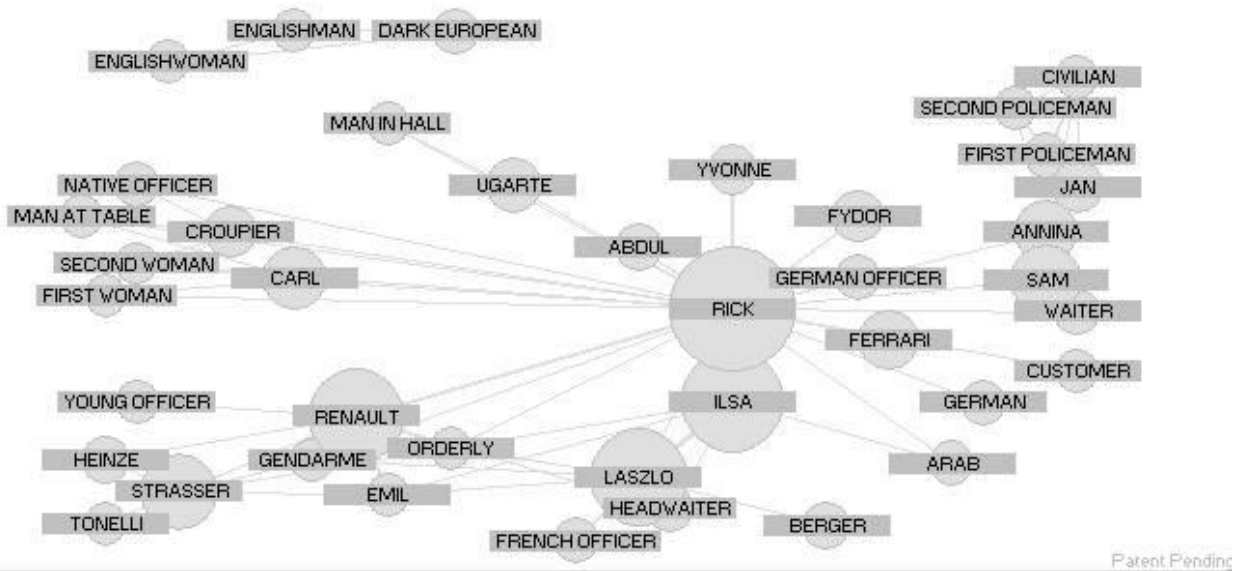


Figure 1.2 - Sophocles Social Network

Patent Pending

The lack of screenplay analytics in screenwriting applications is somewhat surprising, especially considering the range of analytic archetypes represented in the screenwriting 'how-to' books, as discussed in part 2 of this article.