

## WNIM E-marketing Insights 40 – The Long Tail

Written: June 2005

A complete list of all articles written for the E-marketing Insights column of What's New in Marketing is available at: <http://www.davechaffey.com/E-marketing-Insights>.

Visit [www.davechaffey.com](http://www.davechaffey.com) to subscribe to *E-marketing Essentials*, a monthly briefing for marketers of 5 "Must-read" UK/European-focused articles about digital marketing, compiled from the hundreds Dave reads each month.

<b>About the author – Dave Chaffey, BSc, PhD, MCIM</b>	
<b>E-marketing trainer and consultant</b>	Clients include 3M, BP, Barbican, Britvic, The British Council, Computer 2000, HSBC, Intel, Siebel, NCH and Tektronix
<b>Author of 6 best-selling business books</b>	<ul style="list-style-type: none"> <li>• <i>Internet Marketing: Strategy, Implementation and Practice</i></li> <li>• <i>E-business and E-commerce Management</i></li> <li>• <i>Total E-mail Marketing</i></li> </ul>
<b>A marketing 'guru'</b>	Recognised by the CIM as one of 50 gurus who have 'shaped the future of marketing'. I am also proud to have been recognized by the Department of Trade and Industry, NOP World and E-consultancy as one of the "Top 100 hundred people commended by the industry as key influencers and drivers, who have driven the development and growth of e-commerce in the UK over the last ten years."
<b>E-marketing Director</b>	I work for Ripe Design ( <a href="http://www.ripe.co.uk">http://www.ripe.co.uk</a> ) defining the online marketing strategy and implementation plans for our digital marketing clients.
<b>Visiting lecturer at leading UK business schools</b>	<ul style="list-style-type: none"> <li>• University of Cranfield</li> <li>• University of Leeds</li> <li>• University of Warwick</li> </ul>

### Table of contents

<b>Introduction</b> .....	<b>1</b>
<b>What is the Long Tail?</b> .....	<b>2</b>
Zipf's law .....	2
Power Laws and Probability Distributions .....	3
The Long Tail .....	3
Summary - Implications of the Long Tail for marketers .....	4
<b>Digital marketing applications of The Long Tail</b> .....	<b>4</b>
1. The popularity of web sites in a category measured through unique visitors. ....	4
2. The popularity of search terms within a category. ....	5
3. The popularity of content within a web site. ....	6
4. Conversion to outcomes on a site. ....	7
6. The Recency or Frequency of visits to a web site.....	7
7. Response to an e-mail campaign through time. ....	8
8. The popularity of items purchased from an e-retail site.....	8
Weaknesses with Zipf's law .....	8
<b>Next months article</b> .....	<b>9</b>
<b>References and Further reading</b> .....	<b>9</b>
<b>About the author</b> .....	<b>9</b>

### Introduction

If tracking the pulse of the Internet is any indication, over the last 9 months, "The Long Tail" has become an important concept for marketers, and in particular, digital marketers to consider. Since the concept first came to prominence in an October 2004 article by

Wired magazine's editor-in-chief Chris Anderson, references to it have rocketed. Chris now has a lively blog and will publish a book next year. Seth Godin, has referenced it and you can see that it is still referred to a lot in blogs by viewing the frequency with which it appears on the Blogpulse trace ([www.blogpulse.com](http://www.blogpulse.com)) trace. It has also gained traction offline with prominent articles in the Guardian and Economist.

So, there seems to be a lot of fuss about the Long Tail, but what is it and what are its implications?

We start this article looking at how the concept was first introduced by Chris Anderson and in the second part I look at 7 examples of where we see the long tail in online marketing and examine the implications for marketers.

## What is the Long Tail?

The first thing to know about the Long Tail is that it's nothing new; it's an old concept applied in new times. The second thing to know is that its greatest relevance to marketers is that it arises from the variability in human and consumer behaviour.

### Zipf's law

The phenomenon now referred to as the 'Long Tail' was arguably first applied to human behaviour by George Kingsley Zipf, professor of linguistics at Harvard who observed the phenomenon in word usage (See [http://en.wikipedia.org/wiki/Zipf%27s\\_law](http://en.wikipedia.org/wiki/Zipf%27s_law)). He found that if the variation in popularity of different words in a language is considered, there is a systematic pattern in the frequency of usage or popularity.

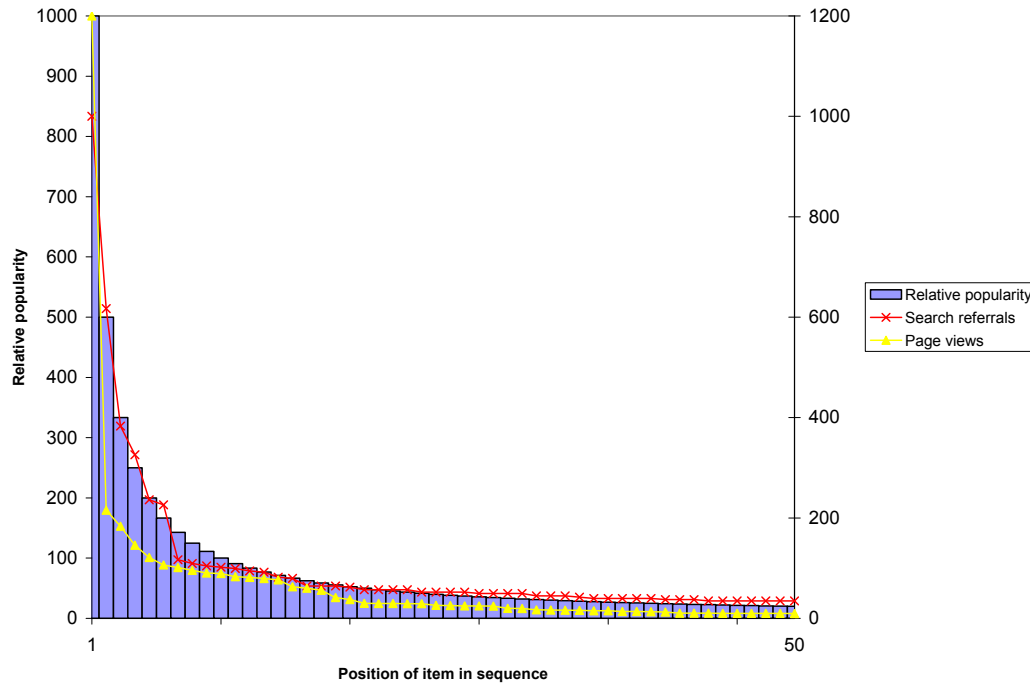
Discussion of The Long Tail shows how Zipf's "law" can be applied to different collections of items which are selected by people. The "Law" shows how the frequency or popularity of the items varies.

Zipf's "law" suggests that if a collection of items is ordered or ranked by popularity, the second item will be around half the popularity of the first one and the third item will be about a third of the popularity of the first one and so on. In general:

*The  $k^{\text{th}}$  item is  $1/k$  the popularity of the first.*

Notice that I say "law" in quotes, since it is not a physical law such as gravitational force, but an observation that there is a tendency to this frequency distribution.

Visuals are essential to understand this law – look at the Figure below which shows how the 'Relative popularity' of items are predicted to decline according to Zipf's law from a maximum count of 1000 for the most popular item to 20 for the 50<sup>th</sup> item.



**Figure – Zipf’s law showing decrease in popularity of items within an ordered sequence. Curves overlaid show variation in relative popularity for search referrals and count of page views from site [www.marketing-online.co.uk](http://www.marketing-online.co.uk)**

## Power Laws and Probability Distributions

The pattern of rapid decline in popularity or frequency of occurrence you can see from the graph is sometimes known as a “Power law”. Such Power laws based on a probability distribution are observed in many fields, including physics, biology, geography, sociology and economics. For example, the size of earthquakes, the sizes of settlements and the level of income all follow a similar pattern.

Pareto’s law or the Pareto distribution, well known to marketers as the “80-20” law, that shows that 80% of our profits come from 20% of our customers is also a Power law. Pareto originally found that 20% of a population owned 80% of the wealth.

## The Long Tail

This Long Tail effect is accentuated online. In Chris Anderson’s original article he refers to the popularity of books or other products stocked by a retailer such as Amazon. He says that Amazon uses its web and its supply chain management approach to create “an infinite shelf”. With the scale of Amazon’s reach into the worldwide online audience, it can still generate profit from the books on the tail since they won’t typically be discounted heavily. Through on-site search and personalisation facilities, a minor audience can be readily connected with the audience that it is interested in them.

But there is more. The Long Tail phenomenon can help drive sales of less popular items when combined with automated personalisation tools such as the “Amazon recommends” feature (readers who bought books on this topic, by one author, may also be interested in another author). Anderson gives the example of the climbing book *Touching the Void* by Joe Simpson. Although you may know this from the recent film, the book is much older, describing an incident in the 1980s, and was originally large known only to climbers. When another similar, but much more popular, mountaineering disaster story about Everest climbers (*Into Thin Air* by Jon Krakauer) was published by 1990s the recommendation software noticed that some people who had bought the Krakauer book had also published

the Simpson book. As a result it promoted the Simpson book thus increasing its sales dramatically, moving it nearer to the head of the Long Tail. Similar phenomena can occur on online DVD rental or music download also.

Anderson says that the implications are that online retailers should make available as wide a range of titles as possible (often on a long-back list) and then fulfil them when demand arises. He also promotes this as a triumph of online economics, saying:

*"For too long we've been suffering the tyranny of lowest-common-denominator fare, subjected to brain-dead summer blockbusters and manufactured pop. Why? Economics. Many of our assumptions about popular taste are actually artifacts of poor supply-and-demand matching - a market response to inefficient distribution."*

## Summary - Implications of the Long Tail for marketers

The Long Tail and Zipf's Law can be applied to describe the variation in preferences for selecting or purchasing from a choice for products as varied as books, CDs, electronic items, travel or financial services. We will see later in this article, that evidence of this variation in preferences is also evident in researching these products online.

Since the tail is long, it is a mistake to concentrate marketing efforts only on the most popular items since many customers or prospects will have a different behaviour and so will have different content or product preferences. At the same time, with limited resources, there are arguments for concentrating marketing efforts on 20% of the product range that will deliver 80% of the volume.

In the context of online retail, Anderson recommends these three rules:

**Rule 1: Make everything available.** Put simply, the more products you make available, the more you will sell! Anderson says that many will say that the long-tail isn't significant – the head is most important according to the 80:20 rule. Andersen argues that the tail is important from a competitive point of view since some competitors will not give their customers access to the full tail and the tail is often more profitable – discounting is not required to compete in the same way it is for the head.

**Rule 2: Cut the price in half. Now lower it.** This refers to the elasticity of pricing. He suggests that the long-tail may enable prices to be reduced to enable increases in volume at greater profitability. His example refers to what he suggests is overpricing for online music services. He gives the example of US provider Rhapsody performing an experiment in elastic pricing where tracks were offered at 99 cents, 79 cents, and 49 cents. Although the 49 cent tracks were only half the price of the 99 cent tracks, Rhapsody sold three times as many of them.

**Rule 3: Help me find it** This can be related back to the Touching the Void example – popular titles have to be available to attract visitors to the site in order for them to become aware of the Long Tail through searches or personal recommendation. The tail should be easy to find through searching – users cannot be expected to find items on the tail by browsing through an extensive hierarchical navigation.

## Digital marketing applications of The Long Tail

In Chris Anderson's original article he largely limited his discussion of the Long Tail to retail products, but any online marketers familiar with web analytics – the study of customer behaviour online through web metrics – will be able to quickly identify other applications.

Chris's article prompted me to think of these examples and implications of online behaviour that follow the Long Tail principle:

### 1. The popularity of web sites in a category measured through unique visitors.

If the number of visitors to all web sites, or sites in a category are plotted in order, they will show a similar pattern to that in **Error! Reference source not found.** This point has

been noted by Jakob Nielsen in 1997 and 2003 (See <http://www.useit.com/alertbox/20030616.html> for details). He explains:

*'Simply stated, big sites get disproportionately more traffic than smaller sites. A site ranked number 100, for example, will get 10 times more traffic than a site ranked number 1,000. (In general, site N gets M/N times the traffic of site M.)*

*Small sites have two huge advantages over big sites: there are many more of them and they are more specialized and thus more targeted. Small sites speak directly to the specific needs and interests of a committed user community, and thus have much higher value per page view.'*

A similar indication of popularity is evident in the number of links to different sites. See for example the number of votes for the top blogs (<http://www.technorati.com/live/top100.html>) and top shared bookmarks on social network Delicious (<http://del.icio.us/popular>).

*Implication:* E-communications techniques such as interactive advertising, affiliate marketing and link-building can be used to take advantage of the Long Tail. Using such techniques to communicate with potential visitors on niche sites can be a relatively low cost approach to achieving reach in comparison to expenditure on the top 10 portals of the web or a category. To deal with the number of sites on which to place online ads or affiliate links, ad and affiliate networks have been created to act as brokers between the site owners and the merchants.

## **2. The popularity of search terms within a category.**

This includes both searches entered into a search engine such as Google and on an individual site (on site search).

The curve in the figure earlier in this article shows referrals to my [www.marketing-online.co.uk](http://www.marketing-online.co.uk) fit the theoretical distribution very closely. Incredibly, the top 100 most popular keyphrases account for only 22% of all search terms that refer visits to the site, showing how long this tail is!

The frequency distribution of referrals from other sites shows a similar pattern (Google much more important, then Yahoo and MSN and then many smaller third party sites).

Another example of the variation in search behaviour, showing all searches across a single category is viewable from this summary by Danny Sullivan on *Search's Long Tail* (<http://blog.searchenginewatch.com/blog/050314-164653>). Danny refers to the "onesies and twosies" – those phrases which may only generate one or two visits per month, but may be collectively important.

Typically the complexity of phrase varies typed varied according to the stage of research about a product or content – initially searchers may enter a two or three word phrase (the popular phrases in the head), followed by a longer more precise phrase as they refine their search (the less popular phrases in the tail).

Data published by Hitwise ([www.hitwise.co.uk](http://www.hitwise.co.uk)) in 2005 for a well-known travel site showed that well over 10,000 different brand and travel-related keyphrase terms are used to attract visitors to this site.

*Implication:* Keyphrase analysis used to determine which search terms to use for Search engine optimisation and Pay Per Click marketing is most effective when hundreds of potential phrases are analysed for each customer need (and on-site outcome) rather than a handful of keyphrases used in some cases.

Different search marketing strategies should be developed to exploit the characteristics of online user behaviour. Search engine marketers talk about strategies to target the head and the tail.

For example, one search engine marketing strategy could involve Search Engine Optimisation (SEO) approach to target the head with a Pay Per Click approach used to target the more specific and generally lower cost phrases in the tail.

Of course this is a simplification and there are exceptions to this approach; SEO strategy can also be used to exploit the tail. Speaking at the E-metrics Summit 2005 in London ([www.emetrics.org](http://www.emetrics.org)), members of Lastminute.com's business intelligence team described how on-page copy for 12,000 product pages from their product database was optimised for search engines in order to connect searchers with less common destination-specific phrases. For example, a page labelled "Self-catering holidays in St.Lucia", will be rare, but if someone does type it, a company such as Lastminute wants to be visible.

Conversely, Pay Per Click marketing could be selectively used to target the Search term "head". For example, typing "car insurance" shows many companies who are unable, in the short-term to be listed in the natural listings, using paid search to target the head.

This discussion raises the question of how many keyphrases a company should review to select which to optimise or advertise on. It is difficult to generalise, since it depends on the variation in behaviour of the audience and the range of products offered. One, well-known European shopping comparison site optimises on over a million keyphrases due to the variation in search behaviour and the range of products it offers! Fortunately, most companies have fewer products. Even so, for a retailer, financial services provider or travel company, the answer is probably 'thousands' and even for companies with a more limited range of products, the answer is 'hundreds'

One way of tackling the 'how many keyphrases' question is to consider the "80:20rule". This suggests that 80% of your search volume will be in the head, within the top 20% of phrases. The number of keyphrases you need to select to be visible or represented within the first one or two search results pages is then indicated by the top 20% of phrases are used within your market. This data can be gleaned through sources such as Wordtracker ([www.wordtracker.com](http://www.wordtracker.com)), Overture ([www.overture.com](http://www.overture.com)) or Hitwise ([www.hitwise.com](http://www.hitwise.com)).

If there are just 100 phrases above a cut-off volume of say 10 searches per month in your market, this would indicate that optimising or advertising on 20 phrases would give you coverage of 80% of potential searches. However, as we have seen above, there are often thousands, or tens of thousands of phrases used in a given market. At a volume of 1,000 search phrases in a market above 10 searches a month, you need to be visible for 200 phrases to give you 80% coverage and at a volume of 10,000 phrases, you need to be visible for 2,000 phrases!

Take care with this simplistic approach though, since as we have said, the value is often in the tail – typically less competitive, lower cost phrases which are more likely to convert. The tail phrases only have this characteristic though in immature markets, it is interesting to speculate on what will happen when more marketers become adept at exploiting the tail. Competition will likely drive up paid search prices such that only those who are efficient at converting their customers and those who can derive the highest lifetime value from their customers will be able to afford to compete! Those companies which succeed will also likely have selected the best bidding strategies and the best bid management software (for example Atlas One Point and Bid Buddy) to manage keyphrases in the tail.

### **3. The popularity of content within a web site.**

The curve shown in the figure for the count of page views of different pages on Dave Chaffey's <http://www.marketing-online.co.uk> site is shown by the curve with triangles. Here there is a less good fit with the theoretical distribution and the tail is less extended with 90% of page views included within the top 50 pages. The tail is less long here since relatively few visitors use the site search on this site and the majority are constrained by the navigation and so are not likely to visit such a wide variety of pages.

*Implication:* the more pages you have with different content relevant to your audience, the more likely you are to meet the needs of your audience and the more they will engage with the content. This helps both for visitors arriving direct at a site and through arriving from search engines.

Having the right navigation, personalisation showing related content or products and on-site search to connect the audience with information about the tail is important also.

The more links there are to content elsewhere on the site will also “reduce the friction” on a page and is more likely connect a diverse audience with its diverse needs. Have you ever considered why the Amazon home page (<http://www.amazon.com>) has so many tabs and links and is a relatively long page? Well, its all to do with the Long Tail – it helps connect Amazon’s audience with its product. The same is also true on Product-specific pages deep within a site – for example an insurer selling its products will do better when it has more links answering questions the customer may have. Compare for example, <http://www.norwichunion.com/single-trip-travel-insurance> which provides more detailed links to other contents than <http://www.rac.co.uk/insurance/travel>).

#### **4. Conversion to outcomes on a site.**

These outcomes can include sale, lead or simply finding the right piece of content.

It is well known that there are high attrition rates on e-retail sites, with only a relatively small proportion (usually less than 10%) completing the transaction.

*Implication:* It is less clear, in this instance, what the relationship to the Long Tail is. But I would suggest that there are a range of barriers that stop the conversion and they will vary in frequency due to the range in consumer preferences. For example, the most common problems are likely to be related to usability or trust about privacy or security, with some other reasons, for example, about copy being less important. Conversion specialist use web analytics, A/B or multivariate tests on new designs, surveys, focus groups and usability studies to determine what these problems are, and then to resolve them. We can suggest it is relatively easy to identify the most frequent problems, but it will be less easy to identify less frequent problems (e.g. problems with copy) and there will be diminishing returns and improvements in conversion through solving these problems.

#### **5. Dwell times on a web site or web page.**

If web analytics software is used to visualise the duration of visits to a site, measured through number of minutes for each visitor session or number of pages viewed, a typical Long Tail distribution will be found. The majority of visitors will sadly only stay for a few seconds or 2 or 3 page views. The tail will be those who browse around a site and are interrupted to go off and view other sites or doing things around the home or office.

*Implication:* Measuring duration by time and particularly average time on the site is a poor way to understand behaviour on the site. The average length of time is skewed by people who leave a page loaded for over 30 minutes which is the standard cut-off time for sessions to end within web analytics software. Much better insights are available through considered the number of pages viewed in a session and overlaying the type of content accessed. This then relates back to conversion barriers, with web marketers needing to work out the roadblocks between different visitor intents and outcomes.

#### **6. The Recency or Frequency of visits to a web site.**

This time the variation in consumer behaviour concerns how often they visit a site or use an online service. There tends to be a range in behaviours according to site type and the Long Tail doesn’t always apply.

This may also show a Long Tail frequency distribution, but it will depend on the type of site as this article shows: <http://www.jimnovo.com/graphs.htm>.

For a portal site such as the BBC, MSN and Yahoo!, a typical Long Tail pattern will be found when plotting the distribution of recency or frequency amongst the sites audience (The first curve in Jim’s article). The majority of visitors will visit regularly – hourly or daily, but some will naturally visit less frequently and will have a lower recency also.



But this pattern will be reversed (an inverse Long Tail) for a destination site such as a retail site (the second curve in Jim's article). Most visitors will visit or buy infrequently – perhaps once every few months rather than every day.

Different pattern will be seen in online services which require a log-in facility such as online banking – this may have a normal distribution, but this can be converted to a Long Tail as deviation from the average.

*Implications:* This type of recency and frequency visualisation is valuable in understanding how successful a company is in obtaining repeated online interactions with its customers. These don't have to be sales interactions, they include:

- Sales transactions
- Service transactions (including log-ins)
- Content transactions (returning to view a particular content type such as news)
- Responses to e-mail (opens and clicks)

Again, average lengths of time quoted for recency and frequency need to be treated with caution because of the length of the tail. But as Jim Novo points out in his 'Drilling Down' E-book it is still useful to set 'hurdle rates' as targets for achieving certain levels of activity within a customer-base. For example, an online bank may seek to achieve 30% of customers servicing their account within a 60 day period.

## **7. Response to an e-mail campaign through time.**

Typically around half of visitors will respond to e-mail by opening or clicking within 48 hours, but there will be a gradual decline over the next month or so as people have time to respond.

We often see in an E-newsletter with many links, that some linked content options will be much more popular than others, similar to that for a web page.

*Implication:* Related content and images need to be maintained on a server for as long as realistic. It should not be assumed in follow-up communications, that everyone has read the e-mail after 2 or 3 days.

More diverse content and offers may encourage higher clickthroughs. For example a single offer and clickthrough message is less likely to appeal to a diverse audience in comparison with 2 or 3 incentives to clickthrough. Of course, a balance has to be struck between offering choice and losing focus of the message and offer.

## **8. The popularity of items purchased from an e-retail site.**

This is the main point in the original Wired piece summarised earlier in this article. There will be some items that are very popular and purchased frequently, but a Long Tail of items purchased less frequently, but significant overall.

*Implication:* A larger inventory will result in more sales. Higher profit margins are possible for less popular items since consumers may be prepared to pay more for difficult-to-obtain items. The Wired article explains how the popular items may draw visitors to an e-retail site, but how recommendation and personalisation tools can make them aware of items on the tail.

## **Weaknesses with Zipf's law**

In the digital world, the law is most accurate when the choice of consumers is not constrained by other factors. – for example keyphrases to search for are not constrained, but pages on a web site or links within an e-mail are constrained and influenced by the navigation and offers on the home page.

While we have suggested that the law is valuable since it highlights a range of customer behaviours and we should accommodate the full range of behaviours this is not always practical or desirable in practice. We often shouldn't concentrate equally on all the different preferences for sites, content or products.



Given that resources are always limited, we still have to prioritise our marketing attention on the most popular items. For example, with search engine marketing it is logical to concentrate efforts on Google, Yahoo! and MSN as the three most popular search engines. Small improvements in search marketing efficiency on these engines can have the biggest returns. Likewise, the search terms which are most important within an industry, for example 'cheap flights' for the airline industry warrant special attention. Creating special pages or PPC ads to attract these visitors is worthwhile, but there a point of diminishing returns which is soon reached where the cost of additional promotional effort on less popular terms is not offset by the returns. The exception to this is where pages or ads can be automatically generated and optimised for search – a trick often used by affiliate marketers.

## Next months article

Next time, we look at more practical matters – with all the reports about cookie deleting and blocking by consumers, can digital marketers trust them any more?

## References and Further reading

Anderson, C. (2004) The Long Tail. *Wired*. 12.10. October 2004.  
<http://www.wired.com/wired/archive/12.10/tail.html>

Chris Anderson now has a blog site ([www.thelongtail.com](http://www.thelongtail.com)), the Long Tail to support a book on the topic to be published in 2006.

Danny Sullivan – Search's Long Tail <http://blog.searchenginewatch.com/blog/050314-164653>

Wikipedia on the Long Tail and Power Distributions  
([http://en.wikipedia.org/wiki/Long\\_Tail](http://en.wikipedia.org/wiki/Long_Tail))

Seth Godin – Thinking about the Long Tail (part 1)  
[http://sethgodin.typepad.com/seths\\_blog/2005/03/thinking\\_about\\_.html](http://sethgodin.typepad.com/seths_blog/2005/03/thinking_about_.html)

## About the author

Dr Dave Chaffey is workshop leader for a range of one-day e-marketing training workshops from the CIM:

- E-mail Marketing ([www.cim.co.uk/0766](http://www.cim.co.uk/0766))
- Running Effective E-marketing Campaigns ([www.cim.co.uk/0767](http://www.cim.co.uk/0767))
- Improving Your Results from Digital Marketing ([www.cim.co.uk/1138](http://www.cim.co.uk/1138))
- Marketing Research Using the Internet ([www.cim.co.uk/1135](http://www.cim.co.uk/1135))

Go to [www.cimtraining.com](http://www.cimtraining.com) for course details and online booking.

Dave Chaffey is trainer and consultant for Marketing Insights Limited ([www.marketing-insights.co.uk](http://www.marketing-insights.co.uk)) and E-marketing Director at Ripe ([www.ripe.co.uk](http://www.ripe.co.uk)). He is a prolific e-business author whose books include '*Total E-mail Marketing*', '*Internet marketing: Strategy, Implementation and Practice*' and '*E-business and E-commerce Management*'.

Read Dave Chaffey's blog ([www.davechaffey.com](http://www.davechaffey.com)) for E-marketing Essentials – the 5 "must-read" articles about online marketing from the hundreds Dave reads each month.