



Stream mining for news analysis

What is the challenge?

The internet provides constant, massive streams of information. Each minute these streams deliver thousands of potentially relevant pieces of information. Such information might be newswire messages, analyst opinions, statements by technology or market research firms, company announcements, blog entries, etc.

Current approaches to this massive flow of information are no longer adequate: **Filtering** (e. g. alerts or search agents) cannot effectively reduce information overload; even more important, it misses „unknown unknowns“: information that is relevant but that the user does not even know might exist. **Searching** only taps information across one particular time slice and does not reveal topic developments over time.

The challenge is to make sense of internet information streams.

How does mergeflow help?

mergeflow is being developed specifically to deal with information streams. Its benefits:

Handle at least ten times more news compared to manual analysis.

mergeflow automatically generates structured views on the news. This helps users to quickly navigate their current information landscape. Users can focus on interpreting the news, rather than on collecting them.

Effectively deal with redundancy. Most events are covered by more than one information source. mergeflow groups together news on the same event so that the user does not have to.

Discover new, relevant topics and trends. The latest news are not always news. They might just revisit a topic that has been reported on before. mergeflow automatically distinguishes „new“ from „latest“. The user does not have to wade through thousands of news manually anymore to find this difference.

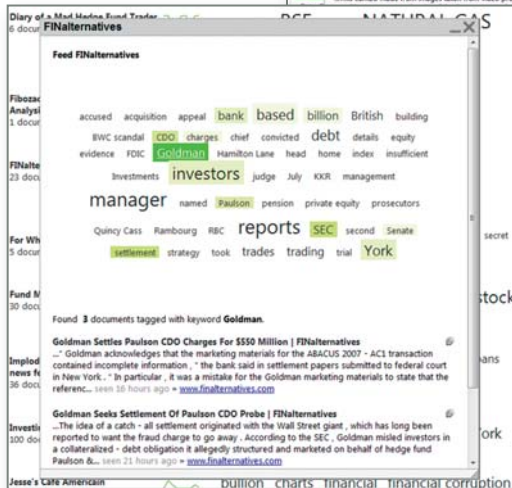
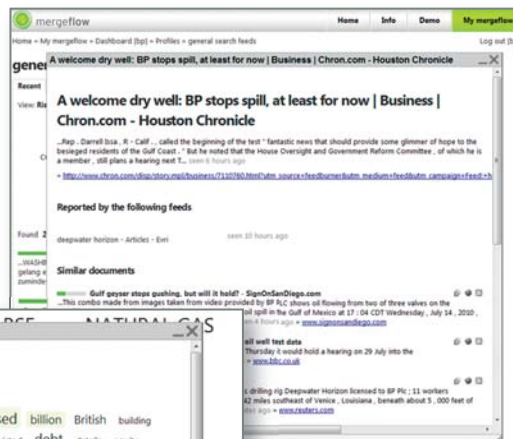
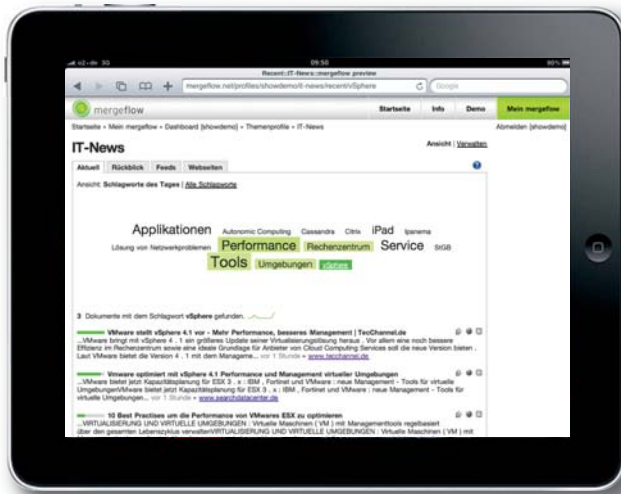
Understand information in its context. Each piece of information has a context that provides an „information ecosystem“. In order to understand a piece of information, one has to consider its information ecosystem. With mergeflow users can grasp this ecosystem. mergeflow shows topically related news, further information from the same source, topic threads through time, or the geographical location of a topic.

How does mergeflow work?

Completely automatic methods. We believe in using completely automatic news analysis algorithms. We want our software to deliver neutral information, not information that someone thinks is important to you.

Flexibility. Virtually any internet news source can be added to mergeflow with just a few mouseclicks. This provides users with the freedom to configure mergeflow to best suit their individual interests.

SaaS (software as a service). mergeflow is completely SaaS-based. Users do not have to install any software. They can use any modern web browser to access mergeflow from anywhere, any time.



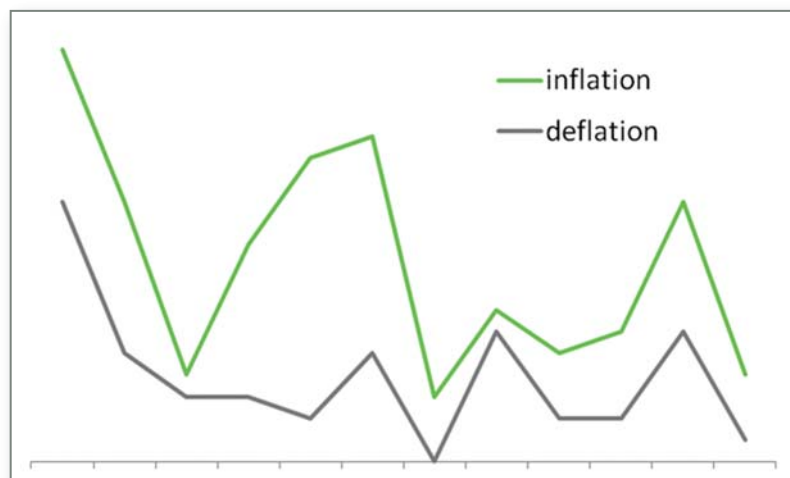
mergeflow works for any topic, any source, and with any browser on any platform (including the iPad).

Examples

Here are a few examples of how mergeflow can be used in financial-services-related news analyses.

General economic topics

Per se, „inflation“, „deflation“ and „double dip“ are abstract economic concepts. Yet, of course, these concepts are mentioned in thousands of news items every day. But what are their concrete, „real world“ contexts? Who talks about double dip (this might indicate whether „double dip“ is special interest or mainstream)? Are topics coming up in connection with „inflation“ the same as those coming up in the context of „deflation“? In order to address this question, mergeflow automatically creates „newness indicators“. These newness indicators show whether information is new or whether it (or something similar) has been seen before. The chart below shows mergeflow newness indicators for inflation and deflation over time (= two weeks in July 2010).



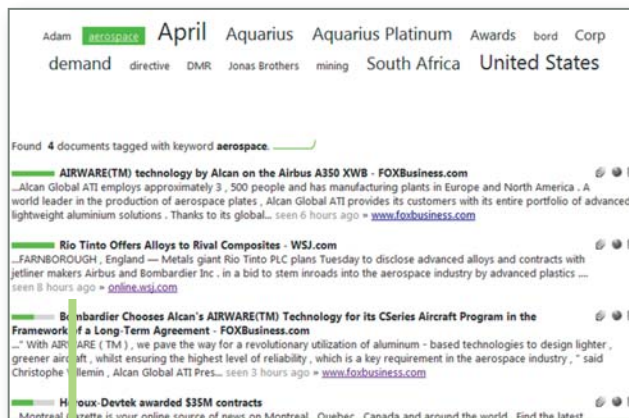
Automatically generated „newness indicator“ for inflation and deflation over time (two weeks in July 2010).

The correlated lines in this chart suggest that whenever new topics come up in the context of inflation, something new comes up for deflation too (and vice versa). A sample of news articles confirms that these new pieces of information are in fact often the same for inflation and deflation: many articles currently (July 2010) discuss whether we will see inflation or deflation. Note that this picture most likely will change at different times in the economic cycle.

Commodities research

Analysts covering commodities have to follow a large number of information streams: general business and finance news (e.g. from newspapers or blogs), research services and consultancies (Brook Hunt, CRU, etc.), statements by other analysts or experts, news on relevant companies, and, of course, news on the commodities. Following all these information streams manually is not realistic; it would mean wading through thousands of news each day.

mergeflow provides a structured view on these information streams. Among other benefits, this helps to spot topics that are relevant but that the user did not know exactly how to search for. In one application, we used mergeflow to structure news on selected commodities (e.g. aluminum, platinum, silver, gold, copper). By automatically determining news and keywords with high „newness value“, mergeflow hinted the user at news on aluminum composites in the aerospace sector. Considering the increasing use of composite materials in aerospace, this might develop into an interesting market.



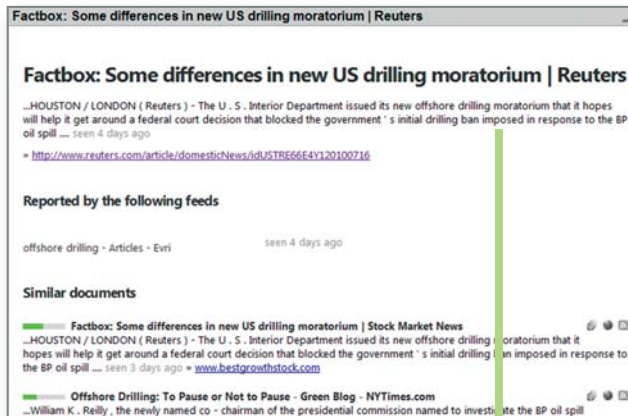
From thousands of news, mergeflow determines keywords with high „newness value“ and alerts the user to potentially interesting new market developments.



Credit research

In order to accurately gauge the creditworthiness of a business, analysts need a 360° view on the business. Particularly for large corporations that appear in the news all the time, this is a daunting task. The analyst has to monitor analyst and executive statements, research reports, general news, company briefings, technical information, legal (regulatory) news that might affect the business, etc.

In one application of mergeflow, the goal was to provide a constant „information baseline“ on BP, particularly also the temporal development of this information baseline. With mergeflow, the user could drill down to relevant new information much quicker: instead of reading more than 1200 news articles that appeared over the course of a few hours, topic clusters automatically generated by mergeflow trimmed down the „information search space“ to 10 relevant clusters of topically related news. One of these clusters contained new, potentially relevant information on the US drilling moratorium:



Company information

Software by mergeflow AG provides effective access to relevant information. From document archives and information streams, our software extracts relevant topics, detects new trends, and reduces information complexity.

mergeflow AG is a member of the Microstep Alliance

Contact information

e-mail	info@mergeflow.com
phone	+49 (0)89 43 7777 94-0
fax	+49 (0)89 43 7777 94-99
internet	www.mergeflow.com
	mergeflow.net
address	mergeflow AG
	Buelowstr. 27
	81679 Muenchen
	Germany

