

# When Machines Dominate Humans: The Challenges of Mining and Consuming Machine- generated Web Mail

Yoelle Maarek  
Yahoo Labs  
MATAM, Advanced Tech Park  
Haifa 31905, ISRAEL  
yoelle@ymail.com

## ABSTRACT

In spite of personal communications moving more and more towards social and mobile, especially with younger generations, email traffic continues to grow. This growth is mostly attributed to (non-spam) machine-generated email, which, against common perception, is often extremely valuable. Indeed, together with monthly newsletters that can easily be ignored, inboxes contain flight itineraries, booking confirmations, receipts or invoices that are critical to many users. In this talk, I will discuss the new nature of consumer email, which is dominated by machine-generated messages of highly heterogeneous forms and value. I will show how the change has not been fully recognized yet by my most email clients (as an example, why should there still be a reply option associated with a message coming from a “do-not-reply@” address?). I will introduce some approaches for large-scale mail mining specifically tailored to machine-generated email. I will conclude by discussing possible applications and research directions.

## Categories and Subject Descriptors

H.4.3 [Information Systems Applications]: Communications Applications, *Electronic Email*

## Keywords

Web Data Mining, Machine Generated Email.