

On the Semantics of Email Addresses

Brian Rabern*

At the turn of the century, electronic mail emerged as a core mode of communication, fundamentally transforming the way individuals and organizations interact. In this new digital age, one ubiquitous representational device is the *email address*. Given the importance of email addresses in our lives, it is natural to ask: what do email addresses represent? And how do they represent it?

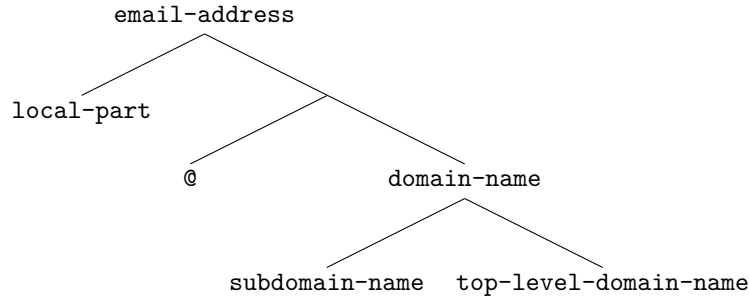
An email address, such as `username@nyu.edu`, is a compound expression that refers to an electronic *mailbox*, where a mailbox is an entity responsible for managing incoming and outgoing messages, typically utilizing a storage location on a computer server. Email addresses have a formal syntax, specified by the Internet Message Format (RFC 5322)¹, that essentially consists of a *local-part* and a *domain-name* separated by an at-sign (@). Given that email addresses have a semantic interpretation, and that they have a formal syntax, it is reasonable to assume that they have a compositional semantics. Do they? If so, what is it?

In order to answer this question, we first need to specify the syntax in slightly more detail. But we needn't go into all the fine details of the RFC, which has to do with which characters are allowed in the basic expressions (e.g., letters, digits, dots, hyphens, etc.). Instead we can give a grammar that captures the important structure of email addresses and blurs over some of the internal lexical details. We can specify the syntax of email addresses in Backus–Naur form as follows:

```
email-address ::= local-part "@" domain-name
local-part   ::= sequence of characters
domain-name  ::= subdomain-name "." top-level-domain-name
subdomain-name ::= sequence of characters
top-level-domain-name ::= sequence of characters
```

*`brian.rabern@graph.fm`, GraphFM/University of Edinburgh: <https://brianrabern.net>

Or more visually we can represent the syntax in a tree diagram:



With the syntax in place we can work out the semantics. Let's start with some basic semantics facts about email addresses:

- Fact 1.** Distinct email addresses can refer to the same mailbox.
- Fact 2.** Distinct mailboxes cannot be referred to by the same email address.
- Fact 3.** Distinct email addresses can refer to different mailboxes, even if they have the same local-part, or if they have same domain-name.
- Fact 4.** Distinct mailboxes can be associated with the same user.
- Fact 5.** Some email addresses don't refer to any mailbox.

We've noted already that an email address refers to an abstract entity, a mailbox. So we know, for example, that:

$$\llbracket \text{username@nyu.edu} \rrbracket = \text{mailbox_123}$$

But what does the domain-name part of the email address, “nyu.edu”, refer to? It refers to a digital *domain*—an abstract object that maps to specific digital resources, which in turn locates the physical server hosting those resources. So, for example, the domain-name “nyu.edu” might refer to a particular domain, call it domain_nyu. So we have:

$$\llbracket \text{nyu.edu} \rrbracket = \text{domain_nyu}$$

If the domain-name refers to a domain, then the subdomain-name refers to something that maps a domain to a domain. For example, the subdomain-name “nyu” might refer to a function that maps the domain “edu” to a particular domain:

$$\llbracket \text{nyu.edu} \rrbracket = \llbracket \text{nyu} \rrbracket(\llbracket \text{edu} \rrbracket) = \text{domain_nyu}$$

Whereas the subdomain-name “**ucla**” might refer to a different function that maps the domain **edu** to a different domain, call it **domain__ucla**:

$$\llbracket \text{ucla.edu} \rrbracket = \llbracket \text{ucla} \rrbracket (\llbracket \text{edu} \rrbracket) = \text{domain_ucla}$$

But what about the local-part? What does that refer to? Given Fact 3, we know it doesn’t refer to a mailbox. Two email addresses with the same local-part can refer to different mailboxes, e.g., **brian.rabern@gmail.com** and **brian.rabern@graph.fm** have the same local-part but refer to different mailboxes.

Perhaps the local-part refers to a person, or a user that the mailbox is associated with. After all, we sometimes use someone’s username to refer to them, e.g., “**dughof**” (the local-part of **dughof@indiana.edu**) might be used to refer to the user Douglas Hofstadter. While it is true that the local-part can be used as a nickname of sorts, it can’t be assumed to refer to the person. This is because of Fact 4: Distinct mailboxes can be associated with the same user. For example, these are two of my email addresses:

brian.rabern@gmail.com

brabern25@gmail.com

The mailboxes that these refer to are both associated with me and they are on the same domain, but they are nevertheless distinct mailboxes. So the assumption that the local-part refers to a person or user, leads to a failure of compositionality. Substituting “**brian.rabern**” for “**brabern25**” in **brabern25@gmail.com** doesn’t preserve reference to the same mailbox.

The local-part doesn’t refer to a user and it doesn’t refer to a mailbox. But we know that the entire email address refers to a mailbox, and the domain-name part refers to a domain. So the local-part would seem to be a kind of *handle* that when combined with the domain via the at-sign, gives us the mailbox. We will call the referent of the local-part, a “handle”—and it may be that distinct local-parts can refer to the same handle. For example, consider subaddressing, where it might be that the local-parts “**brian.rabern**” and “**brian.rabern+logic**” refer to the same handle. Or they might not. We could instead understand a handle to simply be the string that comprises the local-part. We needn’t decide this now.

The at-sign, then, can be understood as referring to a function that maps a domain to a function that maps handles to mailboxes.

$$f_{@} : \text{Domians} \rightarrow (\text{Handles} \rightarrow \text{Mailboxes})$$

To make this more precise, let a model \mathfrak{A} be a quadruple $\langle D, H, M, I \rangle$, where D is a set of digital domains, H is a set of handles, M is a set of mailboxes, and I is an interpretation function which maps the basic expressions to appropriate

entities. For the semantics of the basic expressions we then have the following clauses:

- If h is a local-part, then:

$$\llbracket h \rrbracket_{\mathfrak{A}} = I(h), \text{ where } I(h) \in H$$

- If d is a top-level domain-name, then:

$$\llbracket d \rrbracket_{\mathfrak{A}} = I(d), \text{ where } I(d) \in D$$

- If s is a subdomain-name, then:

$$\llbracket s \rrbracket_{\mathfrak{A}} = I(s), \text{ where } I(s) \in D \rightarrow D$$

- If $@$ is the at-sign, then:

$$\llbracket @ \rrbracket_{\mathfrak{A}} = I(@), \text{ where } I(@) \in D \rightarrow (H \rightarrow M)$$

The referent of any complex expression is, then, determined by the referents of its parts and functional application. If `lp@sub.dom` is an email address, composed of a local-part `lp`, a subdomain `sub`, and a top-level domain `dom`, then (supressing the model subscript \mathfrak{A}):

$$\begin{aligned} \llbracket \text{lp@sub.dom} \rrbracket &= \llbracket @ \rrbracket (\llbracket \text{sub} \rrbracket (\llbracket \text{dom} \rrbracket)) (\llbracket \text{lp} \rrbracket) \\ &= \llbracket @ \rrbracket (\llbracket \text{sub} \rrbracket (d)) (\llbracket \text{lp} \rrbracket) \\ &= \llbracket @ \rrbracket (s(d)) (\llbracket \text{lp} \rrbracket) \\ &= f_{@}(s(d)) (\llbracket \text{lp} \rrbracket) \\ &= f_{@}(s(d)) (h) \\ &= m \end{aligned}$$

The reference to a mailbox is determined by the function $f_{@}$ that maps the domain $s(d)$ to a function that maps the handle h to a mailbox m . In this way, the semantics of email addresses is compositional—the referent of the email address is determined by the referents of its parts and the syntactic structure of the address.

Let's consider an example to illustrate the compositional semantics of email addresses. Consider the following email addresses

`chalmers@nyu.edu`

`dc126@nyu.edu`

These emails refer to the same mailbox.

$$\llbracket \text{chalmers@nyu.edu} \rrbracket_{\mathfrak{A}} = \llbracket \text{dc126@nyu.edu} \rrbracket_{\mathfrak{A}}$$

We can derive this result assuming that the function $\llbracket @ \rrbracket_{\mathfrak{A}} = f_{@}$ is such that it maps the domain-name **nyu.edu** to a function that maps the handles referred to by “**chalmers**” and “**dc126**” to the same mailbox. Let’s assume that the handles are distinct so that

$$\llbracket \text{chalmers} \rrbracket_{\mathfrak{A}} = \text{‘chalmers’} \neq \llbracket \text{dc126} \rrbracket_{\mathfrak{A}} = \text{‘dc126’}$$

Then we have:

$$f_{@}(\llbracket \text{nyu.edu} \rrbracket)(\llbracket \text{chalmers} \rrbracket) = f_{@}(\llbracket \text{nyu.edu} \rrbracket)(\llbracket \text{dc126} \rrbracket)$$

$$f_{@}(\text{domain_nyu})(\text{‘chalmers’}) = f_{@}(\text{domain_nyu})(\text{‘dc126’})$$

$$f_{@}(\text{domain_nyu})(\text{‘chalmers’}) = f_{@}(\text{domain_nyu})(\text{‘dc126’})$$

$$\text{mailbox_djc} = \text{mailbox_djc}$$

What about Fact 5? What should we say about an email addresses such as

$$\text{rabern@hotmail.com},$$

which no longer refers to a mailbox? It’s syntactically well-formed but it doesn’t refer to anything. It’s possible that the domain-name **hotmail.com** no longer refers to a domain, or that the function $f_{@}$ doesn’t map the handle to a mailbox. This is analogous to an improper definite description like “the king of France” which is syntactically well-formed but doesn’t refer to anything. There are a number of standard ways to handle this, but we will leave this open for now.

There are no doubt many other interesting questions that can be asked about the semantics of email addresses. But the aim here was to show that email addresses have a compositional semantics, and to sketch out what that might look like. One set of issues that we haven’t touched on are the meta-physical questions. What is a domain? What is a mailbox? What is a handle? These ontological questions are difficult, but it is not clear that they need to be answered in order to give a compositional semantics for email addresses. After all, it seems that we can give a compositional semantics of numerals without answering the question of what numbers are. Or we can give a semantics for ‘I’ without answering the question of what a person is, or what the *self* is?

Notes

¹See rules for addr-spec in “Internet Message Format”, RFC 5322, October 2008. Available at: <https://tools.ietf.org/html/rfc5322>